The Role of Headhunters in Wage Inequality: It's All about Matching

Alexey Gorn[*]

December 5, 2016

JOB MARKET PAPER

Link to Most Current Version

Abstract

This study relates the observed increase in levels and dispersion of the U.S. top wages to the increasing prominence of headhunters (professional recruiters). I illustrate the main results with a theoretical model that incorporates headhunters in the labor market framework of random search and two-sided heterogeneity. In the model, headhunters improve assortative matching between firms and their top employees via two channels: passive on-the-job search and screening of candidates. The calibrated model shows that headhunters can account for 40% of the increase in the top 1% wage share and 70% of the increase in the top 10% wage share in the U.S. from 1970 to 2010. I provide supporting empirical evidence on the importance of headhunters for the rise in top wages based on cross-country evidence on headhunter hires/fees and top income growth, as well as on micro evidence for CEO compensation in the U.S.

Keywords: wage dispersion, top incomes, sorting, on-the-job search, headhunters

[*] *Bocconi University, e-mail: alexey.gorn@phd.unibocconi.it.* Many thanks to Antonella Trigari, Nicola Gennaioli, Nicola Pavoni, and Yuri Tserlukevich for valuable discussions and advice. Also thanks to Nicolas Serrano-Velarde, Tito Boeri, and the participants of Bocconi PhD Research Seminar with Alberto Alesina, 2014 EDGE Jamboree in Copenhagen, Workshop on Dynamic Macroeconomics in Vigo, and EEA Congress in Geneva. I gratefully acknowledge financing by Fondazione CARIPLO Thesis Research Grant.

1 Introduction

Top wages have been rising sharply in the United States from the early 1970s. One of the reasons for this rise is improved matching between firms and employees at the top positions (especially CEOs and top managers), as documented by Song, Price, Guvenen, Bloom, and von Wachter (2016). The conventional view attributes the improvement in matching to a large extent to skill-biased technological change which raises incentives for firms and workers to be better matched. However, while this explanation is successful in explaining the rise of the upper-middle class, it has difficulties to explain the sharp rise of top wage. The results by Song et al. (2016) reveal a strong non-linearity in the sorting pattern, that is, a disproportional shift of high-skilled workers to high-paying firms in comparison to medium-skilled workers to medium-paying firms, that cannot be generated by standard models of labor markets.

In this study, I develop a model where improved matching at the top is explained by the increasing role of headhunters, or executive search firms, in the labor market. Headhunters, who started to gain market share in the U.S. in the 1970s and now assist to fill more than half of the positions in the top wage segment, enhance matching for two reasons. First, they provide more suitable candidates for the firm because they can screen the candidates better. Second, they induce passive on-the-job search as they contact potential candidates directly, creating opportunities for new matches without active search from workers on the job. To be clear, the headhunters restrict the pool of potential candidates facing firms to only the high-skilled workers, while at the same time, expand the pool of potential candidates to a larger number of those high-skilled workers. These two features guarantee a good match to both the firm and the high-skilled worker, increasing productivity when the match is formed through a headhunter, and therefore allowing wages of such matches to increase. Because executive search firms operate on the top wage segment, these improvements in the matching do not happen (or happen to a lower degree) over the rest of the distribution, leading top wages to increase much more compared to the rest of the distribution. Headhunters generate a strong non-linearity in matching upgrade because they operate exclusively at the very top.

The main indicator of the relative magnitude of top wages used in the literature is the share of total wages that go to the top 1% or top 10% of all employees (from now on, just the top 1% or top 10% wage share). Both shares increased dramatically in the U.S. since the 1970s. The top 1% wage share rose from 5.1% in 1970 to 10.9% in 2010, while the top 10% wage share rose from 25.7% to 34.5%. Figure 1 plots both shares for the U.S. and compares them to the case of France.

The comparison provides suggestive evidence on the role of headhunters for top wages as in France, a country where headhunters started to gain market share only in the 1990s, top wage shares started to rise only in the mid 1990s.

As we said, among the main reasons behind the sharp increase in top wages considered in the literature is improved matching between firms and employees[1]. Song et al. (2016) show using administrative data that improved sorting between workers and firms was a primary factor for the increase in wage dispersion in the U.S. between 1980s and 2000s.[2] Moreover, a closer investigation of their results in terms of the joint distribution of workers and firms by individual fixed effects (a proxy for unobserved skill and productivity) shows that most of the high-skilled workers moved from low- and medium-productivity firms to high-productivity firms over the years. Improvement in the matching is thus concentrated in the top part of the distribution with high-skilled workers toward the end of the sample almost exclusively employed in high-productive firms[3]. This pattern is consistent with the headhunters creating a separate market for high-skilled workers, thus improving matching at the top.

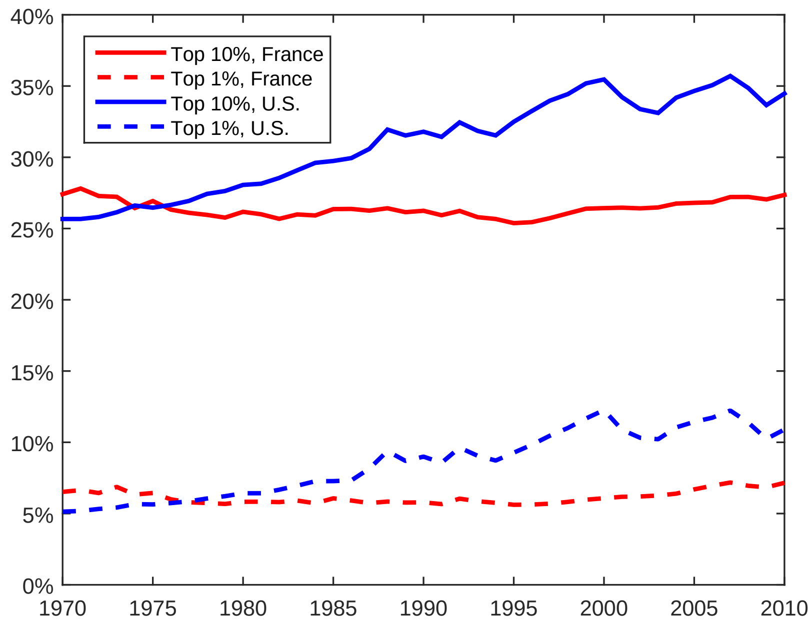
Headhunters were gaining market share in the U.S. at the same time the improvement in sorting at the top was taking place. The fee revenues of top headhunter companies increased by a factor of 10 from the 1970s to 2010s. The International Association of Corporate and Professional Recruiters (IACPR) report (2003) states that 54% of the positions paying above $150,000 a year (around top 5% wages) in 2003 were filled through headhunters, which constitutes a significant

[1] Alternative explanations include: decrease in top income taxes - Alvaredo, Atkinson, Piketty, and Saez (2013); direct effects of skill-biased technological change on wages - Acemoglu (2002), Katz, Kearney, et al. (2006), and many others; social norms - Piketty (2014); exogenous changes in random growth theories - Gabaix, Lasry, Lions, and Moll (in press), Jones (2015), Jones and Kim (2014), and Aoki and Nirei (2015); and numerous studies on the increase of CEO pay who constitute a significant part of top 10%, including Gabaix and Landier (2008), Bell and Reenen (2013), and Lemieux, MacLeod, Parent, et al. (2009) among others.

[2] In line with the result of Song et al. (2016), Card, Heining, and Kline (2013) find that the increased covariance between workers' and firms' fixed effects contributed to 30% of the increase in wage dispersion in West Germany between 1985 and 2009. Bagger, Sørensen, and Vejlin (2013) document similar findings using Danish data. They find that the correlation between worker and firm fixed effects increased from -0.07 in 1981 to 0.14 in 2001. Then splitting the sample by the quartiles of worker fixed effects, they find that the aggregate effect was driven by the top quartile of workers where the correlation increased from -0.20 to 0.37, while the correlation stayed almost unchanged at around zero for the rest of the quartiles. Håkanson, Lindqvist, and Vlachos (2015) also document significant increase in sorting in Sweden between 1986 and 2008.

[3] Estimating a fixed effect regression is only one way to evaluate sorting in the labor market. Eeckhout and Kircher (2011) show potential problems with identifying sorting with estimated fixed effects. Other studies propose non-parametric techniques to assess the degree of sorting. In general, such studies find a higher degree of sorting than in the studies using fixed effects regressions. Notable examples are Hagedorn, Law, and Manovskii (2012) who find the correlation between worker and firm ranks to be 0.75 in Germany, and Lise, Meghir, and Robin (2016) who find significant sorting for college-graduates in the U.S. These studies mainly focus on the methodology and identifying sorting in a particular period, and not studying the change in sorting over time. An exception is the study by Schulz and Lochner (2016) who show, using non-parametric techniques, that sorting increased in Germany between 1998 and 2008.

Figure 1: Top 10% and 1% shares of the total wage bill in the U.S. and France.



This figure plots the top 10% share of the total wage bill in the U.S. (solid blue) and France (solid red), and the top 1% share of the total wage bill in the U.S. (dashed blue) and France (dashed red). Source: Piketty (2014).

share of the labor market, especially for top-paying positions. Importantly, headhunters are not only focused on specific industries or positions but rather cover a wide range of positions across all industries in the economy. A detailed description of the structure and workings of the headhunter industry, as well as the way it has evolved over time, will be presented in the next section. Their key attribute is that they improve matching between firms and employees as they aim at finding the best candidate for a position. Individual headhunters are typically focused on a specific position or industry and build detailed databases with information on the majority of potential candidates for such position or industry. With this detailed information already in hand, when asked to assist to fill a position they can choose the best fitting candidate and improve matching. This is the first main feature of headhunters - better screening of the candidates. After an headhunter chooses a candidate from its database, it calls the candidate directly and asks whether she wants to consider a job offer (without specifying the offer). The headhunter contacts any candidate who is perceived to be the best fit for the position without the candidate having to signal interest in changing job. A worker who has not to put effort into receiving an offer from an headhunter and who agrees to consider the offer is essentially searching passively on the job. Passive on-the-job search is the second major feature of the headhunter industry. Passive search helps high-skilled workers not getting stuck for long in positions not fitting them, moving them to a better fitting position and improving aggregate matching.

To quantify the contribution of better matching induced by headhunters to the increase in top wages, I develop a labor market model along the lines of Mortensen and Pissarides (1994) augmented with heterogeneous workers and firms[4]. I introduce the headhunter industry in the model by adding a new channel for matching workers and firms. Firms with an open position can either post a vacancy as in the standard model or hire through an headhunter. The difference for the firm is that while it cannot screen workers coming through vacancies, the headhunter guarantees the firm a minimal skill level of the worker with whom the firm is matched. This because only workers with skills above a certain threshold can receive call from headhunters. Consider now the worker's side. Low-skilled workers have access to the standard channel to be matched with a firm with an open position, and to search on the job they have to pay the search costs. I call this active on-the-job search. For high-skilled workers, instead, on top of active on-the-job search, there is also a possibility of passive on-the-job search. A worker is searching passively, if she agrees to consider an offer when a headhunter calls[5]. Screening and passive on-the-job search are exactly the two main features of the headhunter industry that I introduce to the theoretical model.

Having set up the model, I apply the following calibration strategy. First, I calibrate the model without headhunters targeting moments of the wage distribution and aggregate labor market moments in the 1970s in the U.S.. The key calibrated parameters include those characterizing the exogenous distributions of workers over skills and firms over productivity. The idea is that the U.S. labor market in the 1970s is well approximated by one with no, or limited, role for headhunters. Having fixed the parameters not related to the headhunters, I then introduce the headhunter channel to the model and calibrate the related parameters to target the moments of the headhunter industry in the 2010s. On top of this, I introduce skill-biased technological change to match the increase in 90/50 wage ratio from 1970 to 2010 in order to compare the relative contribution of headhunters and skill-biased technological change for the rise of the top wage shares. Skill-biased technological change is modeled as an increase in the degree of complementarity between workers and firms.

[4]Other studies including two-sided heterogeneity in the labor market include Shimer and Smith (2000), Postel-Vinay and Robin (2002), Teulings and Gautier (2004), Gautier and Teulings (2015) and many others.

[5]Cappelli and Hamori (2013) show that more than half of executives are willing to consider an offer when a headhunter calls them.

To evaluate the role of the headhunter channel, I then exploit the richness of the model and compare several statistics related to the wage distribution with and without the headhunters. Importantly, I perform experiments in line with Song et al. (2016) where I compare results from the model generated data to the U.S. data to see if the improvement in matching in the model has similar features to the observed improvement in matching. My calibration strategy answers the question how would the distribution of wages (and therefore also the top wages) have changed between the 1970s and 2010s if skill-biased technological change and headhunters had been the only factors rising top wage inequality. To assess the relative contribution of the two factors, I then shut down one channel at a time: the increase of the degree of complementarity, or the headhunter channel. Finally, I also give a chance to skill-biased technological change to explain all the increase in wage inequality without the headhunters. To do that, I change the increase in the degree of complementarity to match the increase in the top 10% wage share and assess how the model fits other moments. In the baseline calibration, I use a simple sharing rule for wage setting. The reason for this is that I want to isolate the effects of matching. Headhunters may also affect the bargaining power of firms or employees, for example shifting up wages of employees hired through headhunters. I want to abstract from such effects in the main experiment because it would be difficult to disentangle the relative contributions of matching and wage setting in the change of the wage distribution[6].

The main quantitative result of the paper is that the rise of headhunters accounts for 40% of the increase in top 1% and 70% of the increase in top 10% wage shares in the U.S. from the 1970s to 2010s. Skill-biased technological change contributes to another 10% of the top 1% wage share and 23% of the top 10% wage share increase, and interaction between the two factors rises the top shares by further 10% for top 1% and 5% for top 10%. Headhunters and skill-biased technological change also contribute equally to the rise of the upper-middle class measured as 90/50 wage ratio. Thus, the sharp increase of top wages in the model is mainly due to improved assortative matching after introducing headhunters. With headhunters, high-productive firms employ almost exclusively high-skilled workers and high-skilled workers work almost exclusively in high-productive firms. Comparing joint distributions of worker-firm matches in the two steady states reveals a pattern similar to the empirical results of Song et al. (2016), where almost all types of firms, except the highest-paying, lose the highest-skilled workers and where the highest-paying firms gain those workers disproportionally. The headhunter channel generates the strong non-linearity in the change in assortative matching observed in the data, with

[6] However, I redo the experiments with wages set through explicit bargaining in the appendix.

disproportionate improvement in matching for high-skilled workers. I am not aware of other theoretical models able to generate such non-linearity. If I allow the model to match the increase in the top wages without the headhunter channel, the model overshoots the 90/50 wage ratio by 82%. This happens exactly because of the absence of a strong non-linearity (SBTC affects more high-productive workers, but it is not enough), increase in complementarity shifts the whole distribution of wages to the right almost uniformly. The non-linearity generated by the headhunters allows to shift the right tail of the distribution further apart from the rest of the distribution without changing the overall shape of the distribution in the middle.[7]

The model is a significant improvement to other theoretical models showing the importance of assortative matching for wage distribution. Bagger and Lentz (2014) is the closest study. They show that on-the-job search is a crucial mechanism to generate assortative matching in a Diamond-Mortensen-Pissarides model with two-sided heterogeneity. Bagger and Lentz (2014) consider only active on-the-job search, and this paper shows that passive on-the-job search is crucial for generating even more assortative matching, concentrated at the top, consistent with empirical studies. Uren and Virag (2011), instead, show that skill requirements are important to generate an increase in between-group inequality (increased differences between wages of workers with different skill level). Skill requirements play a similar role as the screening by headhunters. While this paper focusing on headhunters and the top part of the distribution, Uren and Virag (2011) study the overall shape of the wage distribution.

Headhunters play a major role at the top paying segment of the labor market but were overlooked by economic research. This paper is first to bring attention to the phenomenon of headhunters in a macroeconomic perspective and to discuss the effects of their activity on the outcomes of the labor market and the economy as a whole. Introducing headhunters in the theoretical models is another contribution of this paper. Particularly as this allows me to introduce the notion of passive on-the-job search within search models. Passive on-the-job search may prove to be a useful notion also applied to the medium wage segment for which contingent headhunters hire medium paying professionals, to job offers received through referrals, or to direct contacts by employers through websites like LinkedIn. Besides labor markets applications, passive on-the-job search might capture phenomena in search models of financial markets where

[7] The rise of headhunters can be also viewed as a reason for the shift in the mean income growth rates for high-skilled workers in the model of Gabaix et al. (in press). Gabaix et al. (in press) introduce an exogenous increase in mean income growth rate for some workers in 1980s, motivated by globalization and technological change. Headhunters, who allow high-skilled workers to work for the top firms, generate the increase in income growth rate for high-skilled workers due to better matches.

brokers call investors directly offering to buy a particular asset that would fit their portfolio or other theoretical settings.

To provide support for the relevance of headhunters, this paper presents two blocks of independent empirical evidence. First, it uses cross-country differences in the use of headhunters in Europe in 1997 to show that in countries where headhunters made relatively more placements top income shares increased more in the following years. This evidence is in line with the prediction of the model that the more the high-productive firms use headhunters, the better is the improvement of matching at the top, and the higher are the top wages. And second, the paper studies micro data on CEO compensation of listed companies in the U.S.. Main results are that firms pay significantly more to new CEOs comparing to the previous ones, and this difference is higher in the periods when headhunters are used more intensively and in the states where there are less legal obstacles to the activity of headhunters[8]. These results suggest that headhunters indeed improve matching between firms and CEOs and therefore increase the wages at the top. Of course this is just a suggestive evidence, ideally, it should be identified which CEO was hired through a headhunter and which CEO was hired through other sources, but such data is not available, partly because of privacy issues.

On the empirical side, the contribution of this paper is to the hires by headhunters in a country (or region) as a proxy for the degree of assortative matching at the top. It is an empirical challenge to identify sorting in the data. Looking through the lens of the model presented in this paper, the hires by headhunters allow to measure the degree of assortative matching. The more hires at the top are done by the headhunters, the better is the assortative matching at the top.

The remainder of the paper is organized as follows. Section 2 discusses the available empirical evidence about headhunters and headhunter industry. Section 3 presents the theoretical model. Section 4 presents the calibration of the model and discusses the implications of the model on wage inequality. Section 5 provides empirical evidence of cross-country differences in the patterns of top income shares over the last thirty years and their relation to the headhunter industry. Section 6 presents empirical results of the effects of changing CEO on the compensation paid by the firm. Section 7 concludes.

[8]The legal obstacles are instrumented by the enforceability of non-compete agreements as proposed by Garmaise (2009).

2 Headhunter industry

This section offers a short description of different types of headhunters, the practices of headhunters, and the size and the development of the industry over time.

First of all, it is important to distinguish between two types of headhunters - retained headhunters and contingency headhunters. Retained headhunters have an exclusive and often permanent contract with the firm to find and hire an employee for a particular position or several employees for a set of positions. Many retained headhunters have long-term contracts with companies and are used when the firm needs to feel a certain position. Retained headhunters operate on the very top positions in the firm structure that pay above $150,000 a year. Retained headhunters are in the focus of this paper. Contingency headhunters, instead, don't have an exclusive contract with the firm, so sometimes the firm contacts several contingency headhunters and only the one that provides the contact with the candidate that is hired is paid by the firm. Contingency headhunters cover mainly the medium skilled positions such as professionals and general managers, the positions paying between $15,000 and $150,000 a year. Even though contingency headhunters also cover a significant share of the labor market they are left aside in this project because their effect on the top wages is thought to be minor.

In order to operate, the headhunters need to establish a database of potential candidates. The databases are often very detailed and include information about contacts, education, prior experience, particular skills, languages spoken, and many others. The headhunters constantly work on expanding and updating the database. When the headhunter needs to fill a position, the candidate is chosen from the database with a goal to have the best fit for the position. The headhunter then contacts the candidate making an offer to consider a new job possibility, without specifying details about the job (employer, wage range e.t.c.). If the candidate agrees to consider the offer the headhunter interviews the candidate to determine the fit for the position. If the interview is successful, the offer is revealed and the candidate is connected with the firm. Now the firm assesses whether it wants to hire the candidate and if it confirms the candidate is hired by the firm.

Because the costs of creating databases of potential candidates are huge, there are different types of headhunter companies. There are many boutiques with few employees that focus on a particular position in a particular industry (or even sub-industry) and a particular region (sometimes even a city). Because of specialization boutiques are able to find the best candidate for the position of specialization. Then, there are local or specialized companies, who target a specific industry or a region. Finally, there are few international companies who can provide service for almost any top position in any industry and region. In general, headhunters cover a wide range of positions: CEOs, board directors, CFOs, senior executives, general management, top professionals in finance and control, information systems, marketing, and sales. They are not focused only on CEOs as sometimes is perceived but cover almost all the top positions in companies. Industry composition of headhunter operations is also dispersed, they operate in all industries. Distribution of fee revenues by industry in the fourth quarter of 2015 is presented in table 1.

Table 1: Fee revenues of headhunters by industry, 4th quarter 2015.

Industrial	23.5%
Technology	16.2%
Financial	21.0%
Consumer products	18.0%
Life Sciences/Healthcare	15.2%
Non-Profit	4.4%
Other	1.7%

This table presents the distribution of the estimated worldwide fee revenues over industries in the 4th quarter 2015. Source: AESC Insights Q4 2015 State of the Executive Search Industry.

Geographical distribution of fee revenues instead is not so homogeneous, as presented in table 2. Most of the revenues headhunters receive from North America, and mainly the U.S., and Europe is lagging behind with major part of European revenues coming from the U.K. There might be several reasons for the fact that headhunters are more widely used in the U.S. than in Europe. One major reason is the difference in labor market legislation. It is more difficult to be an intermediary in European labor market than in the U.S., especially in some countries. Second possible reason is the cost of creating a database of potential candidates in a new country. The headhunter must know the specifics of the labor market and the companies operating in the country in order to understand the skills demanded by companies and the value of observable signals, such as particular diplomas and experience in particular companies. Because headhunters first appeared in the U.S. they established the databases and acquired the knowledge of the labor market and the signals earlier there and therefore are more efficient.

The history of the rise of headhunters started in the U.S. already in the 1950s with first professional recruiters. However, the first couple of decades were not very successful for them. Only in the late 1970s and early 1980s, the industry started to expand sharply with worldwide fee revenues rising from \$0.75 billion in 1978 to \$3.9 billion in 1990. The fee revenues kept rising up to \$12 billion in 2015 (figure 2a). Partly this rise was mechanical because top wages were

Table 2: Fee revenues of headhunters by region, 4th quarter 2015.

North America	45.5%
EMEA	33.2%
Asia Pacific	16.5%
Latin America	4.8%

This table presents the distribution of the estimated worldwide fee revenues over regions in the 4th quarter 2015. (EMEA - Europe, Middle East, and Africa) Source: AESC Insights Q4 2015 State of the Executive Search Industry.

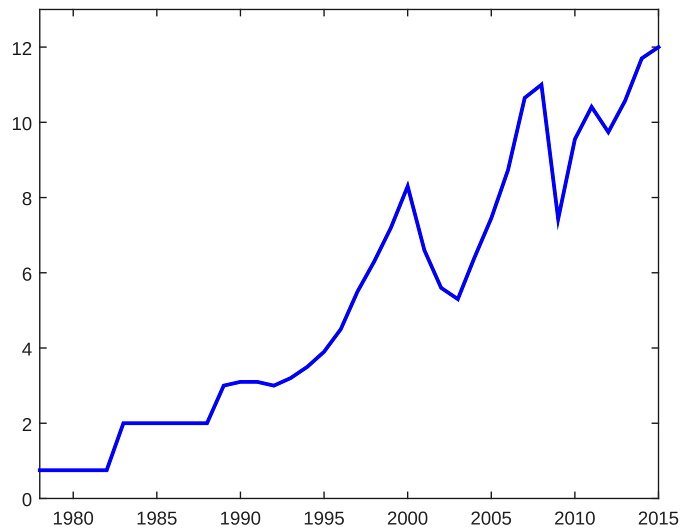
also rising over the same period (figure 2b), but the revenues increase was much larger in proportion to the increase of the wages. Another indicator of the expansion of the industry is the number of hires by headhunters. For example, the number of assignments of one of the historical leaders of the industry Korn/Ferry increased from 42 in 1969 to 8,480 in 2015 according to the financial statements.

There might be several reasons for the rise of headhunters. Maybe the most important reason is the technological progress in IT and communication that increased the quality of the supply of services by headhunters. Developments and growing availability of computers reduced dramatically the costs of managing and searching through databases of potential candidates. Communication technology (mobile phones and emails), instead, made it easier to contact potential candidates and allowed headhunters to expand the networks of potential candidates. Companies that adopted new technologies earlier were more successful in the market[9]. Another effect of technology goes through the demand side; better technology made it easier to apply for jobs (especially in the late 1990s and 2000s). More applications increased the amount of information that the firms had to evaluate to hire a worker and the higher was the position the more information was there to evaluate. It became more efficient for the firms to delegate the screening of workers for top positions to intermediaries - the headhunters - and the demand for headhunter service increased. One more potential reason for the rise that goes through demand is related to the nature of the skills required from employees in the top positions. Because of technological change, globalization, or change of company structure, it became more important for the firm to hire employees with higher general skill in comparison to 1970s. Firms started to use headhunters more because in the 1990s the skill of the CEO, for example, affected the performance of the company much more than in 1970s, while the nature and efficiency of

[9]Jenn (1995) writes: "The drive towards a more consistent quality of service throughout the world has been greatly assisted by the application of information technology to the search business and the use of global databases. Technological advances have allowed firms to search more widely and communicate more efficiently. Virtually all executive search firms are attempting to modernise their communications and database systems on a global basis. ... This is the area where the search world is changing most dramatically. Firms have a tremendous opportunity to improve their efficiency, achieve better margins and differentiate themselves from their competitors."

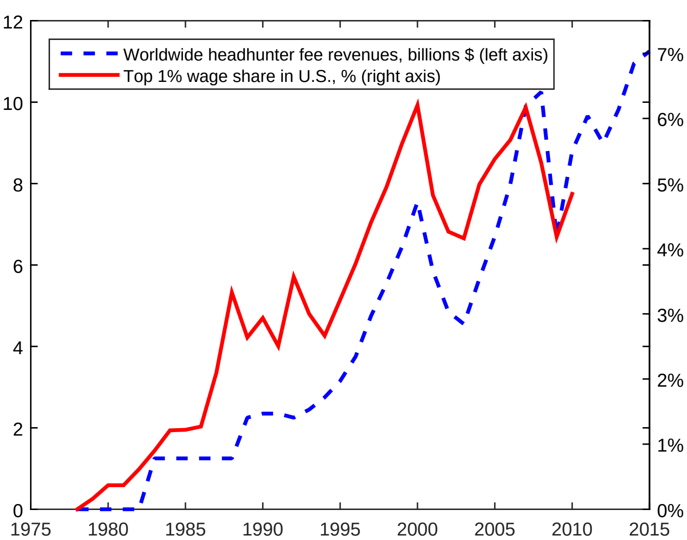
Figure 2: Estimated worldwide fee revenues of headhunters and top wages.

(a) Estimated worldwide fee revenues of headhunters.



This figure plots the estimated worldwide fee revenues of the headhunter industry from 1978 to 2015, in $ billions. Source: AESC Insights Q4 2015 State of the Executive Search Industry.

(b) Cumulative change in headhunters fee revenues and top 1% wage share.



This figure plots the cumulative change in the estimated worldwide fee revenues of the headhunter industry from 1978 to 2013 (dashed blue line, left axis), in $ billions, and the cumulative change in the top 1% wage share in the U.S. from 1978 to 2010 (solid red line, right axis), in %. Both series are normalized to 0 in 1978. Sources: AESC Insights Q4 2015 State of the Executive Search Industry, Piketty (2014) , and author's calculations.

headhunters stayed the same. Even though there is more evidence of the technological supply story, this paper doesn't exclude other reasons for the rise of headhunters. Because of the nature of the experiments done in this paper, the actual reason for the rise doesn't play a big role for the results of the paper. It will be important, however, to understand the exact reason for the rise of headhunters better in the future studies of dynamics of inequality, and the question of the rise of the industry is interesting by itself.

To determine the exact market share of the very closed and private industry is a very difficult task. One can use IACPR report (2003) that shows that 54% of the positions above $150,000 a year were filled by headhunters in 2003. Another way to determine the market share of the headhunters is to compare the estimates of the fee revenues to the ones implied by the total wage bill. Total fee revenues in the U.S. estimated by AESC are around $4.6 billion. From the other side, it is possible to compute the total wages that go to the top 5% of the U.S. employees from the total wage bill and the top 5% wage share from studies on inequality. Then, using the hiring rate one can determine total wage that goes to the new hires in a given year. From the total wages paid to the new hires, headhunters receive as a fee around 30% of the first year wage. It is possible to determine what would be the aggregate fee revenues if the share of the market of headhunters is known. Using the average for the U.S. hiring rate of 3.5%, to be consistent with the estimates by AESC, the share of the headhunters must be around 15% in the labor market for positions in the top 5%. However, the hiring rate at the top is, in general, much lower than in the lower-paying jobs, with tenures being significantly longer. With more realistic hiring rate, the implied market share of headhunters is more than 30%. This estimate is very imprecise because exact average fee paid to headhunters and exact hiring rate are unknown.

3 The model

3.1 Environment

The economy is populated by a continuum of heterogeneous workers differing in their skill level, e, who supply inelastically one unit of labor if employed. When a worker is unemployed she benefits from home production activity, unemployment subsidies, leisure and other possible sources she cannot enjoy during employment, collectively represented by b. Also, there is a continuum of heterogeneous firms that differ in their productivity level, p. Each firm can hire one worker. To do this a firm needs to post a vacancy or to contact a headhunter company. Both workers and firms discount their future utility with discounting rate β.

All workers, unemployed and employed, can search for a job. Each period workers decide whether to search for a job checking vacancies (search actively) and/or to be available for a headhunter company (search passively) if her skill is higher than a threshold $\hat{e}$. Workers searching for a job and firms posting a vacancy are matched randomly by a standard CRS matching technology. Firms using headhunters are randomly matched with workers with a skill level above $\hat{e}$ with a possibly different matching technology. In the baseline model the wage in a match is determined period by period as a fraction of resulting production of the match. The production of a match depends on the firm's productivity level and the worker's skill level. Firms can choose only one channel while workers can search both actively and passively (if they are eligible) at the same time. Separation of matches depends on two factors: idiosyncratic exogenous separation shock, s; and worker quitting to another job, $s_Q(.)$. There is no aggregate uncertainty in the model. The paper considers only stationary equilibria.

3.2 Timing

The time is discrete. Inside every period, first, existing matches produce and wages and unemployment benefits are paid. Then exogenous separations happen. Workers decide in which markets to participate, new firms decide to enter the market and choose the market to search. After that, workers searching for a job and firms searching for a worker are matched.

3.3 Matching

There are two channels in the labor market: the vacancy, V, and the headhunter, H, channels. In channel $i \in \{V, H\}$ workers and firms meet by a standard matching technology: $m_i = m_i(u_i + a_i, j_i)$, where m_i is the number of matches, u_i and a_i are the numbers of unemployed and employed workers participating in the channel, respectively, and $j_i \in \{v, h\}$ is the number of firms participating in the channel. The job finding rate for a worker using channel i is $f_i(u_i, a_i, j_i) = \frac{m_i(u_i+a_i,j_i)}{u_i+a_i}$ and the firm's worker finding rate is $q_i(u_i, a_i, j_i) = \frac{m_i(u_i+a_i,j_i)}{j_i}$.

3.4 Wages and production technology

In the baseline model, the wage is proportional to the match productivity[10]: $w(e,p) = \psi \cdot y(e,p)$ with $0 < \psi < 1$. Where $y(e,p)$ is increasing and quasi-concave in both components, that is $y'_p > 0$, $y'_e > 0$, $y''_{pp} \leq 0$ and $y''_{ee} \leq 0$. Moreover, $y(e,p)$ has a property of complementarity between the worker's skill and firm's productivity, having positive cross-derivatives: $y''_{ep} > 0$, $y''_{pe} > 0$. The production function must be supermodular so that it generates the incentives for positive assortative matching.

3.5 Worker problem

Consider first the problem of a low-skilled unemployed worker. The *low-skilled unemployed worker* is excluded from the headhunter channel, so the only choice that she does is between searching through vacancies and not searching. Low-skilled worker's value of search can be written as:

$$S_U(e) = \max\{S_{UV}(e), 0\}, \; if \; e < \hat{e}, \tag{1}$$

where:

$$S_{UV}(e) \equiv f_V(.)\, E_{p|V}\left[\max\{W(e,p), U(e)\} - U(e)\right] - c_{wV}.$$

If she decides to search through vacancies, with probability $f_V(.)$ she will receive an offer that will be drawn from a distribution of firms posting vacancies. She also has to pay the search cost, c_{wV}, every period of actively searching. When the worker receives an offer she will choose between working in a firm with productivity p and getting the difference between the value of employment, $W(e,p)$, and unemployment, $U(e)$, or staying unemployed and having no gain.

For the *high-skilled unemployed worker*, the problem is the same but she chooses between four options: to search through vacancies, to wait for a headhunter call, to do both, or be inactive. The value of searching through vacancies is the same as for the low-skilled unemployed worker. The value of waiting for a headhunter call, or searching passively, differs in four respects: it has different probability of an offer arrival, $f_H(.)$; it has different search cost, c_{wH}; the search cost is paid only if the offer arrives; and, the offer is drawn from a different distribution - distribution

[10]Other wage setting mechanisms and their implications for the model and main results are considered in the appendix.

of firms using headhunters. The value of searching through both channels is just a combination of the two, with an implicit assumption that better firms are using the headhunter channel[11]. The value of the search of a high-skilled unemployed worker can be written as:

$$S_U(e) = \max\{S_{UV}(e), S_{UH}(e), S_{UVH}(e), 0\},\ if\ e \geq \hat{e}, \tag{2}$$

where

$$S_{UH}(e) \equiv f_H(.)\left(E_{p|H}\left[\max\{W(e,p), U(e)\} - U(e)\right] - c_{wH}\right),$$

$$\begin{aligned} S_{UVH}(e) \equiv \;& f_H(.)\left(E_{p|H}\left[\max\{W(e,p), U(e)\} - U(e)\right] - c_{wH}\right) + \\ & + f_V(.)\,(1 - f_H(.))\,E_{p|V}\left[\max\{W(e,p), U(e)\} - U(e)\right] - c_{wV}. \end{aligned}$$

The unemployed worker who did not receive or reject an offer this period consumes the unemployment benefits and continues the search in the next period. The value of unemployment for a worker can be written as:

$$U(e) = b + \beta\left(U(e) + S_U(e)\right). \tag{3}$$

Now consider an *employed worker*. She also decides whether to participate in the markets, but with a different outside option. Similarly to the unemployed workers, the employed low-skilled worker may choose between searching through vacancies or not to search at all. Now, the outside option for the worker is her previous employment rather than unemployment. For a *low-skilled employed worker* the value of search can be written as:

$$S_E(e,p) = \max\{S_{EV}(e,p), 0\},\ if\ e < \hat{e}, \tag{4}$$

where

$$S_{EV}(e,p) \equiv f_V(.)\,E_{p'|V}\left[\max\{W(e,p'), W(e,p)\} - W(e,p)\right] - c_{wV}.$$

The employed high-skilled worker may choose again between four options: searching through vacancy, headhunter, or both channels, or not to search at all. For a *high-skilled employed worker* the value of search can be written as:

[11] This assumption is used already here for the convenience of notation.

$$S_E(e,p) = \max\left\{S_{EV}(ep), S_{EH}(e,p), S_{EVH}(e,p), 0\right\}, \; if \; e \geq \hat{e}, \tag{5}$$

where the value of search through the vacancies is exactly the same as for a low-skilled worker and

$$S_{EH}(e,p) \equiv f_H(.)\left(E_{p'|H}\left[\max\left\{W(e,p'), W(e,p)\right\} - W(e,p)\right] - c_{wH}\right),$$

$$\begin{aligned} S_{EVH}(e,p) \equiv \;& f_H(.)\left(E_{p|H}\left[\max\left\{W(e,p'), W(e,p)\right\} - W(e,p)\right] - c_{wH}\right) + \\ &+ f_V(.)\left(1 - f_H(.)\right) E_{p|V}\left[\max\left\{W(e,p'), W(e,p)\right\} - W(e,p)\right] - c_{wV}. \end{aligned}$$

If a worker decides to stay in a firm or does not receive an offer, she consumes the current wage this period, and next period the match can be exogenously separated with probability s, in which case the worker becomes unemployed, or with probability $(1-s)$ the match survives and the worker can continue to search on-the-job. The value of work for any worker is:

$$W(e,p) = w(e,p) + \beta\left(sU(e) + (1-s)\left(W(e,p) + S_E(e,p)\right)\right). \tag{6}$$

3.6 Firm problem

Firms also choose channels in the same manner as the workers, but they all solve the same problem (regardless of their productivity level) and they may choose only one channel, having the value of a vacant job defined as:

$$V(p) = \max\left\{V_V(p); V_H(p); 0\right\}. \tag{7}$$

If the firm decides to post a vacancy, it pays the per-period cost $c_{fV} \cdot p$ and is matched with a worker with probability $q_V(.)$ and hires her if she accepts the offer. It happens when the worker doesn't have a better offer at the same period and if she worked in a worse firm (if searching on-the-job). The value of posting a vacancy for a firm is:

$$V_V(p) = -c_{fV} \cdot p + \beta\left(V(p) + q_V(.)\, E_{e|V}\left[P(A)\left(J(p,e) - V(p)\right)\right]\right), \tag{8}$$

where $P(A)$ is the probability of acceptance of the offer by the worker. In general, this probability might depend on many factors: the skill of the worker, the productivity of this firm, the productivity of the current employer of the worker, another offer to the worker, etc. But in the structure of equilibrium that will be considered here, the exact functional form for this probability is relatively simple.

Similarly, if a firm decides to hire through a headhunter, it pays a per-period cost $c_{fH} \cdot p$ and is matched with a worker with probability $q_H(.)$. The value of hiring through a headhunter is:

$$V_H(p) = -c_{fH} \cdot p + \beta \Big(V(p) + q_H(.) \, E_{e|H} \left[P(A) \left(J(p,e) - V(p) \right) \right] \Big). \tag{9}$$

If the firm hires a worker or a match stays for this period, the firm receives the product of the match, pays the wage and in the next period the match is separated for exogenous reasons with probability s or because the worker quits to another firm with probability $s_Q(.)$[12], and the firm has to substitute the worker, or survives otherwise. So the value of a firm can be written as:

$$\begin{aligned} J(p,e) = \;& y(e,p) - w(e,p) + \\ & + \beta \left(\left(s + s_Q(.) (1-s) \right) V(p) + \left(1 - s_Q(.) \right) (1-s) \, J(p,e) \right). \end{aligned} \tag{10}$$

The free entry condition for the firms is the following:

$$E_p \left[V(p) \right] = F, \tag{11}$$

where F is a fixed cost of creating a firm that is paid once to enter the market. It is assumed that before entering the market, firms do not know their level of productivity. Another interpretation of this assumption can be that a firm is created with not just a single vacancy but with a distribution of positions that it needs to fill.

3.7 Steady-state separating equilibrium

In this section, a particular structure of the equilibrium will be considered. The most reasonable scenario, given supermodularity of the production function, is that high-productive firms

[12] The arguments of the function are dropped for convenience of notation, but in general it depends on the firm productivity level, worker's skill level, and characteristics of the labor market in both channels.

would hire through the headhunter channel while low-productive firms would use the standard channel.

Distributions

First, we need to specify distributions that will be used in expectations. Let $F(p)$ be an initial distribution of firm productivity levels and $G(p)$ the measure of firms with an open vacancy (both CDFs have support $\left[\underline{p}, \overline{p}\right]$). Denote as $\hat{p}$ the cutoff level of firm productivity, such that firms with productivity above $\hat{p}$ hire through a headhunter and firms below $\hat{p}$ post a vacancy, so the fraction of firms posting a vacancy is $\frac{G(\hat{p})}{G(\overline{p})}$.

Let $H(e)$ be the initial distribution of all workers over skill levels, $L_V(e)$ be the measure of employed workers searching for a job through the vacancy channel, $L_H(e)$ the measure of employed workers searching for a job through the headhunter channel, $L_{VH}(e)$ the measure of employed workers searching for a job through both channels, and $U(e)$ the measure of unemployed workers over the skill level (all with support$[\underline{e}, \overline{e}]$).

Finally, let $\Phi(e,p)$ be the joint measure of active matches. And $\Lambda_i(e,p)$ be the measure of active matches in which a worker is searching for a new job through channel $i \in \{V, H, VH\}$.

Workers

Given the structure of the equilibrium under consideration and the distributions defined above we can now specify the expectations.

Under our assumptions, low-skilled workers are excluded from the headhunter channel so they can search only through vacancies. Their value of search will be:

$$S_U(e) = S_{UV}(e) \equiv f_V(u_V, a_V, v) \int_{\underline{p}}^{\hat{p}} (W(e,p) - U(e))\, dG(p) - c_{wV}. \tag{12}$$

For high-skilled unemployed workers it is optimal to search through both channels, so their value of search is the following:

$$\begin{aligned} S_U(e) = S_{UVH}(e) \equiv\ & f_H(u_H, a_H, h) \left(\int_{\hat{p}}^{\overline{p}} (W(e,p) - U(e))\, dG(p) - c_{wH} \right) + \\ & + f_V(u_V, a_V, v)\,(1 - f_H(u_H, a_H, h)) \int_{\underline{p}}^{\hat{p}} (W(e,p) - U(e))\, dG(p) - c_{wV}. \end{aligned} \tag{13}$$

Under our assumption, better firms use the headhunter channel, so if unemployed high-skilled worker receives an offer from a headhunter she will accept it regardless of receiving an offer through vacancy channel or not. Instead, this worker will accept an offer from vacancy channel only if she doesn't receive an offer from the headhunter channel at that period.

Again, for employed workers the value of search is the same as the value of search for unemployed except from the outside option. We can define the value of search of low-skilled employed worker as:

$$S_{EV}(e,p) \equiv f_V(u_V, a_V, v) \int_{\underline{p}}^{\hat{p}} \max\left\{W(e,p') - W(e,p);0\right\} dG(p') - c_{wV}. \tag{14}$$

Or, opening the max inside the integral:

$$S_{EV}(e,p) \equiv f_V(u_V, a_V, v) \int_{p}^{\hat{p}} \left(W(e,p') - W(e,p)\right) dG(p') - c_{wV}.$$

The employed worker accepts a new match only if she is matched with a better firm.

To search from employment, the value of search for a worker with skill e working in a firm with productivity p must be positive:

$$S_{EV}(e,p) \geq 0.$$

This equation (when satisfied with equality) implicitly determines the level of the firm productivity such that a worker with a skill level e does not search for a new job: $\tilde{p}_V(e)$ (for $e < \hat{e}$). So that, if a worker with skill e works in a firm with productivity below $\tilde{p}_V(e)$, she searches for another job and doesn't search otherwise. Her value function of searching will be as before:

$$S_E(e,p) = \max\left\{S_{EV}(e,p);0\right\}.$$

For a high-skilled employed worker the value of search is now consists of four options, but in this structure of equilibrium, one of them (searching only through vacancies) will never be optimal. So that the value of search can be defined as:

$$S_E(e,p) = \max\left\{S_{EV}(e,p);S_{EH}(e,p);S_{EVH}(e,p);0\right\} = \max\left\{S_{EH}(e,p);S_{EVH}(e,p);0\right\}.$$

For a high-skilled worker with skill level e, there are now two cutoff productivity levels $\tilde{p}_{VH}(e)$ and $\tilde{p}_H(e)$, with $\tilde{p}_H(e) \geq \tilde{p}_{VH}(e)$. If the worker is employed in a firm with productivity below $\tilde{p}_{VH}(e)$ she will search for another job through both channels. If she works in a firm with productivity level between $\tilde{p}_{VH}(e)$ and $\tilde{p}_H(e)$, she will search only through headhunter channel. While if she works in a firm with productivity above $\tilde{p}_H(e)$, she will not search for another job at all. Before defining the conditions that determine these cutoffs we need to define the value functions.

Value of search through headhunter channel for a high-skilled worker can be defined as:

$$S_{EH}(e,p) \equiv f_H(u_H, a_H, h)\left(\int_{\max\{\hat{p};p\}}^{\overline{p}} \left(W(e,p') - W(e,p)\right) dG(p') - c_{wH}\right). \tag{15}$$

And the value of search through both channel can now be defined as:

$$\begin{aligned} S_{EVH}(e,p) \equiv\ & f_H(u_H, a_H, h)\left(\int_{\hat{p}}^{\overline{p}} \left(W(e,p') - W(e,p)\right) dG(p') - c_{wV}\right) + \\ & + f_V(u_V, a_V, v)\left(1 - f_H(u_H, a_H, h)\right) \int_p^{\hat{p}} \left(W(e,p') - W(e,p)\right) dG(p') - c_{wV}, \end{aligned} \tag{16}$$

note that in this case the first integral starts always in $\hat{p}$ because it will never be optimal to search through both channels if a worker is already working in a firm that hires through the headhunter channel. This is the feature of the structure of equilibrium. In other structure of equilibrium, that will be discussed in extensions, not all firms above the cutoff $\hat{p}$ will be using the headhunter channel, so the limits of the integral will be different.

It is easy to see that, given e, $S_{EVH}(e,p)$ is higher than $S_{EH}(e,p)$ for small p, but $S_{EVH}(e,p)$ decreases faster, so they will always have just one intercept. And the equality:

$$S_{EVH}(e,p) = S_{EH}(e,p)$$

defines the cutoff productivity level of searching through both channels for each worker type, $\tilde{p}_{VH}(e)$, while the equality

$$S_{EH}(e,p) = 0$$

defines the cutoff productivity level for searching only through headhunters, $\tilde{p}_H(e)$.

The value functions of working and unemployment are defined as before.

Firms

We can also rewrite the values of posting a vacancy and hiring through headhunter channel given distributions defined above. The value function of firms posting a vacancy in this case is:

$$
\begin{aligned}
V_V(p) = \quad & -c_{fV}\cdot p+\beta\Bigg(V(p)+q_V(u_V,a_V,v)\Bigg(\tfrac{u_V}{u_V+a_V}\textstyle\int_{\underline{e}}^{\hat{e}}(J(p,e)-V(p))\,dU(e)+\\
& +\tfrac{u_V}{u_V+a_V}(1-f_H(u_H,a_H,h))\textstyle\int_{\hat{e}}^{\overline{e}}(J(p,e)-V(p))\,dU(e)+\\
& +\tfrac{a_V}{u_V+a_V}\textstyle\int_{\underline{e}}^{\hat{e}}\tfrac{\Lambda_V(e,p)}{\Lambda_V(e,\overline{p})}(J(p,e)-V(p))\,dL_V(e)+\\
& +\tfrac{a_V}{u_V+a_V}(1-f_H(u_H,a_H,h))\textstyle\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}(J(p,e)-V(p))\,dL_{VH}(e)\Bigg)\Bigg),
\end{aligned}
$$

where the first part in the summation is the expected value of a match after meeting an unemployed low-skilled worker, the second - an unemployed high-skilled worker, the third - an employed low-skilled worker, and the forth - an employed high-skilled worker.

Similarly, the value function of firms using headhunters is:

$$
\begin{aligned}
V_H(p) = \quad & -c_{fH}\cdot p+\beta\Bigg(V(p)+q_H(u_H,a_H,h)\Bigg(\tfrac{u_H}{u_H+a_H}\textstyle\int_{\hat{e}}^{\overline{e}}(J(p,e)-V(p))\,dU(e)+\\
& +\tfrac{a_H}{u_H+a_H}\textstyle\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_H(e,p)}{\Lambda_H(e,\overline{p})}(J(p,e)-V(p))\,dL_H(e)\\
& +\tfrac{a_H}{u_H+a_H}\textstyle\int_{\hat{e}}^{\overline{e}}\tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}(J(p,e)-V(p))\,dL_{VH}(e)\Bigg)\Bigg),
\end{aligned}
$$

where again the first part in the summation is the expected value of a match after meeting an unemployed worker, the second - an employed worker searching only through the headhunter channel, and the third - an employed worker searching through both channels.

One can show that under reasonable conditions on the value of the search costs and production function, the value of hiring through a headhunter, $V_H(p)$, is lower than the value of posting a vacancy, $V_V(p)$, for small p. But, $V_H(p)$ is increasing faster with p. There will be only one intercept between $V_H(p)$ and $V_V(p)$, $\hat{p}$, such that

$$
\max\{V_V(p);V_H(p)\}=V_V(p)
$$

for $p<\hat{p}$ and

$$
\max\{V_V(p);V_H(p)\}=V_H(p)
$$

for $p>\hat{p}$. And the cutoff productivity is determined by

$$V_V(\hat{p}) = V_H(\hat{p}).$$

Now, given the distributions, we can also specify the quit rate of a worker with skill e from a firm with productivity p:

$$s_Q(e,p,\omega) = \begin{cases} f_V(u_V,a_V,v)\left(\frac{G(\hat{p})-G(p)}{G(\hat{p})-G(\underline{p})}\right) & if\ p < \tilde{p}_V(e)\ and\ e < \underline{e} \\ f_H(u_H,a_H,h)\left(\frac{G(\overline{p})-G(p)}{G(\overline{p})-G(\hat{p})}\right) & if\ \tilde{p}_{VH}(e) < p < \tilde{p}_H(e)\ and\ e \geq \underline{e} \\ f_H(u_H,a_H,h)\left(\frac{G(\overline{p})-G(p)}{G(\overline{p})-G(\hat{p})}\right) + & if\ p < \tilde{p}_{VH}(e)\ and\ e \geq \underline{e} \\ +\left(1-f_H(u_H,a_H,h)\right)\cdot & \\ \cdot f_V(u_V,a_V,v)\left(\frac{G(\hat{p})-G(p)}{G(\hat{p})-G(\underline{p})}\right) & \\ 0 & otherwise, \end{cases}$$

where $\omega = (u_V, a_V, v, u_H, a_H, h)$ is a vector of labor market characteristics.

Finally, the value of an active match for a firm is defined as before, substituting the value functions defined above and the function for quit rates.

Aggregation

The aggregates that enter the matching functions are determined as follows. The number of unemployed workers searching through the vacancy channel:

$$u_V = \int_{\underline{e}}^{\overline{e}} 1dU(e).$$

The number of unemployed workers searching through the headhunter channel:

$$u_H = \int_{\hat{e}}^{\overline{e}} 1dU(e).$$

The number of employed workers searching through the vacancy channel:

$$a_V = \int_{\underline{e}}^{\hat{e}} \int_{\underline{p}}^{\overline{p}} 1d\Lambda_V(e,p) + \int_{\hat{e}}^{\overline{e}} \int_{\underline{p}}^{\overline{p}} 1d\Lambda_{VH}(e,p).$$

The number of employed workers searching through the headhunter channel:

$$a_H = \int_{\hat{e}}^{\overline{e}} \int_{\underline{p}}^{\overline{p}} 1 d\Lambda_{VH}\left(e, p\right) + \int_{\hat{e}}^{\overline{e}} \int_{\underline{p}}^{\overline{p}} 1 d\Lambda_{H}\left(e, p\right).$$

The number of firms using the vacancy channel:

$$v = \int_{\underline{p}}^{\hat{p}} 1 dG\left(p\right).$$

And the number of firms using the headhunter channel:

$$h = \int_{\hat{p}}^{\overline{p}} 1 dG\left(p\right).$$

3.8 Equilibrium

The steady state equilibrium, given the initial distributions of workers over skills and firms over productivity, the skill threshold, the matching functions, and the production function, is defined by the value functions, the endogenous distributions, and the decision rules. The decision rules, together with the endogenous distributions, must satisfy the balances such that the endogenous distributions are stationary.

The balances guarantee that the equilibrium distribution is stationary over time. In the equilibrium the inflow of workers to every worker-firm distribution bin must be equal to the outflow of workers from that bin. For a pair of workers with skill e and firms with productivity p, the equilibrium density of active matches, $\phi\left(e, p\right)$, must satisfy:

$$\phi\left(e, p\right)\left(s + s_Q\left(e, p\right)\left(1 - s\right)\right) = i_E\left(e, p\right) + i_U\left(e, p\right),$$

where the left-hand side is the total outflow from active matches (exogenous separations plus endogenous quits), and the right-hand side is the total inflow into the matches from employment, $i_E\left(e, p\right)$, and unemployment, $i_U\left(e, p\right)$, respectively. The inflow from unemployment can be written as:

$$i_U(e,p) = \begin{cases} f_V(u_V, a_V, v)\,\frac{g(p)}{v}u(e) & \text{if } e < \hat{e},\ p < \hat{p} \\ f_H(u_H, a_H, h)\,\frac{g(p)}{h}u(e) & \text{if } e \geq \hat{e},\ p \geq \hat{p} \\ (1 - f_H(u_H, a_H, h))\, f_V(u_V, a_V, v)\,\frac{g(p)}{v}u(e) & \text{if } e \geq \hat{e},\ p < \hat{p} \\ 0 & \text{otherwise,} \end{cases}$$

where the first condition is satisfied when a worker searches only through vacancies and firm posts a vacancy; the second condition is satisfied when a worker searches through headhunters or both and a firm hires through a headhunter; and the third condition is satisfied when a worker searches through both channels and a firm posts a vacancy. Similarly, the inflow from employment can be written as:

$$i_E(e,p) = \begin{cases} f_V(u_V, a_V, v)\,\frac{g(p)}{v}\int_{\underline{p}}^{\min\{p,\tilde{p}_V(e)\}}\phi(e,p')\,dp' & \text{if } e < \hat{e},\ p < \hat{p} \\ f_H(u_H, a_H, h)\,\frac{g(p)}{h}\int_{\underline{p}}^{\min\{p,\tilde{p}_H(e)\}}\phi(e,p')\,dp' & \text{if } e \geq \hat{e},\ p \geq \hat{p} \\ (1 - f_H(u_H, a_H, h))\, f_V(u_V, a_V, v)\,\frac{g(p)}{v}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')\,dp' & \text{if } e \geq \hat{e},\ p < \hat{p} \\ 0 & \text{otherwise.} \end{cases}$$

3.9 Solution method

The steady state equilibrium of the model is solved for using Matlab. To find an equilibrium the following algorithm is used:

1. Guess the search decision rules for workers and firms (the cutoff levels of productivity)
2. State the system of equations corresponding to the balances of endogenous distribution of matched firms and workers
3. Solve the system using non-linear solution methods (trust-region algorithm or Broydn algorithm) for the invariant distribution
4. Compute the value functions given the distributions
5. Compute the decision rules given the value functions
6. Compare the decision rules, if different - update the decision rules and go to 2.

3.10 Extensions

The most important extension of the model is introduced to capture the fact that not all firms hire employees for top positions through headhunters. In the baseline model every firm above the threshold $\hat{p}$ hires through the headhunters. To avoid it, I introduce an additional idiosyncratic non-monetary cost for hiring through a headhunter. Every firm with an open position draws a cost c_{fN} every period that is associated with the headhunter channel. This cost might reflect corporate practice, an existence of a preferred candidate inside the firm, or specificity of the position. Firms with high cost will have to post a vacancy even if they would benefit from a match through a headhunter otherwise. Updated value of an open position, $\tilde{V}(p)$, can be written as:

$$\tilde{V}(p) = \max\left\{V_V(p); V_H(p) - c_{fN}\right\}.$$

This extension doesn't alter the model significantly, only making the value functions and the balances more cumbersome[13], but brings the model closer to the data. Because the proportion of firms using headhunters by productivity is unobservable, in the baseline calibration the distribution of the costs will be chosen such that the firm has the same chance to hire through the headhunters regardless of the productivity.

Other important extensions include different wage setting mechanisms (wage bargaining and wage Bertrand competition between firms), and explicit modeling of headhunter's problem (choice of the threshold and the cost for the firm). These extensions are secondary to the main experiments in the paper and therefore will be discussed in the appendix.

4 Inequality

4.1 Calibration

The calibration strategy is the following. First, the version of the model without the headhunter channel is calibrated to match the labor market in the U.S. in 1970. Then the parameters related to the headhunter channel are calibrated to match the properties of the headhunter industry in the U.S. in the 2010s. To take into account the skill-biased technological change from 1970

[13] Equations for the value functions and the balances are presented in the appendix.

to 2010 I also change the degree of complementarity in the production function to match the increase in 90/50 wage ratio.

To calibrate the model we need to specify the exogenous distributions of workers, $H(e)$, and firms, $F(p)$, the productivity function, $y(e,p)$, the matching function, $m(u,v)$, the skill threshold, $\hat{e}$, the search costs, c_{fH}, c_{fV}, c_{wH}, c_{wV}, and the distribution of the non-monetary cost c_{fN}. The functional form for the initial distributions of firms and workers is chosen to be beta with the same parameters, λ_1 and λ_2, for workers and firms, and truncated on $\overline{p} = \overline{e}$. The matching function has standard Cobb-Douglas form:

$$m(u,v) = Mu^{\sigma}v^{1-\sigma},$$

and the production function has a form:

$$y(e,p) = (e \cdot p)^{\gamma},$$

with normalization $\gamma = 1$ in 1970. It is easy to see that this production function is supermodular with $\frac{\partial^2 y(e,p)}{\partial e \partial p} = \frac{\partial^2 y(e,p)}{\partial p \partial e} = 1 > 0$.

All parameters, except for the search costs in the headhunter channel and the skill threshold are calibrated to match the wage distribution and other labor market variables in 1970. There are seven parameters to calibrate for the steady state without the headhunters: the parameters of the distribution, λ_1 and λ_2, the maximum type, $\overline{p}$ or $\overline{e}$, the separation rate, s, the matching function efficiency, M, and the search cost through vacancy channel for workers, c_{wV}, and for firms, c_{fV}. To set these seven parameters, seven targets are chosen: the top 1% and 10% wage shares, the 90/50 wage ratio, the unemployment rate, the job finding rate, the quit rate, and the estimate of vacancy cost relative to annual worker's wage. Seven parameters are jointly calibrated to match the targets.

For the headhunter channel, there are just four parameters to calibrate - the search costs for workers, c_{wH}, and firms, c_{fH}, the skill threshold, $\hat{e}$, and the share of firms using headhunters, χ[14]. Four targets are chosen - the estimate of the positive response rate by managers to a call by a headhunter[15], the average fee of headhunters, the range of positions filled by headhunters, and the share of firms that use the headhunters for hiring. This way, the targets do not directly affect the wage distribution, so the calibration strategy does not drive the results. Robustness checks

[14] It is equivalent to calibrating a distribution function for the non-monetary cost c_{fN}.

[15] Estimated by Cappelli and Hamori (2013)

for the choice of parameters for the headhunter channel will be presented in the subsequent sections. On top of the headhunter channel I also increase the degree of complementarity, γ, in order to match the change in the 90/50 wage ratio between 1970 and 2010.

The results of the calibration are presented in Table 3. The model matches well the main characteristics of the wage distribution and the labor market in 1970.

Table 3: Calibrated parameters.

Parameter	Value	Target	Data	Model
Wage distribution, 1970				
Beta parameter, λ_1	1.5	Top 1% wage share	5.1%	4.38%
Beta parameter, λ_2	15	Top 10% wage share	25.7%	25.56%
Maximum types, $\overline{p}$, $\overline{e}$	14	90/50 wage ratio	1.91	2.07
Labor market, 1970				
Separation rate, s	0.026	Unemployment rate	5%	4.9%
Matching function, M	0.9	Job finding rate	50%	49.53%
Search cost - vacancies, c_{wV}	1.05	Quit rate	2%	1.87%
Vacancy cost, c_{fV}	0.17	Vacancy cost estimates	8%	8%
Headhunter industry, 2010				
Headhunter search cost, c_{wH}	1.2	Positive response rate	50%	50.52%
Headhunter firm cost, c_{fH}	2.55	Headhunter average fee	30%	30%
Screening threshold, $\hat{e}$	2.98	Range of positions	top 5%	5.05%
Share of firms using headhunters, χ	0.54	Share of firms	$\sim$54%	54%
Skill-biased technological change, 1970-2010				
Degree of complementarity, γ	1.10	Δ 90/50 wage ratio	0.39	0.39

This table presents the result of the baseline calibration.

4.2 Results

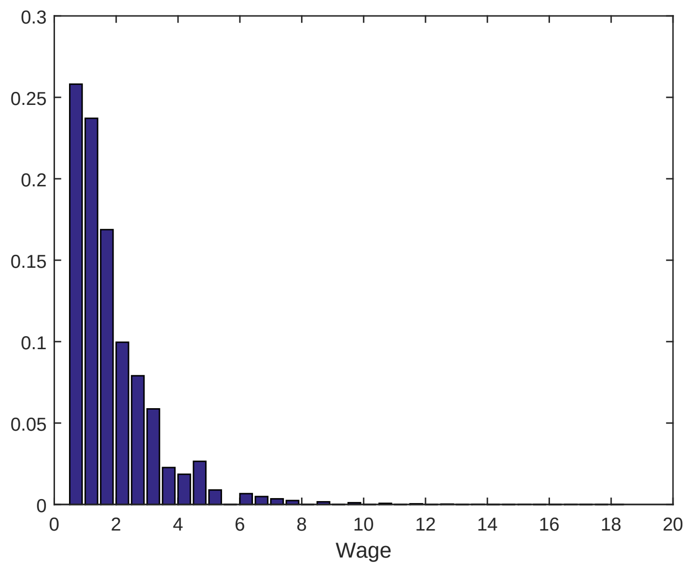
The headhunter channel allows separating high-skilled workers reducing frictions for them and providing them with an exclusive opportunity to work in high-productive firms. The presence of this channel changes the distribution of the workers over the wages. Without the headhunter channel, the distribution has a peak close to minimal possible wage and then decreases, having a form close to Pareto (Figure 3a). When the headhunter channel is present in the model, the distribution still has similar form but has a fatter right tail (Figure 3b), similar to what is observed in the data. The headhunter channel generates the fat tail of the wage distribution in this model. The reason for this is the following, without headhunter channel the probability of matching a high-skilled worker with a high-productive firm is lower than matching a high-skilled worker with a low-productive firm (due to the fact that there are relatively few high-productive firms), so there will be big shares of high-skilled workers working in low-productive firms and low-skilled workers in high-productive firms. Because of skill complementarities, wages of low-skilled workers are lower than wages of high-skilled workers in the same type of firm. And because only some high-productive firms will be matched with high-skilled workers there will be a small mass of workers getting very high wages. When, instead, there is a possibility to hire only high-skilled workers through the headhunter channel, high-productive firms will be matched only with high-skilled workers and all of them will receive relatively high wages; this corresponds to the fat tail of the distribution. The difference between the distributions (Figure 3c) clear indicates the appearance of fatter right tail with the headhunter channel. However, there are two effects changing the wage distribution in this case - headhunters and the skill-biased technological change. Skill-biased technological change increases wages of all workers and therefore moves the whole distribution to the right.

To see the effect of only the headhunter channel on the wage distribution, Figure 4 plots the difference between the distributions without the effects of the skill-biased technological change. An interesting observation about the effect of headhunters on wage distribution can be done - the headhunter channel generates an effect similar to job polarization , namely, decrease in the number of medium-paying jobs and increase in the number of high- and low-paying jobs. This effect also comes from the fact that low-skilled workers move from high-productive to low-productive firms (from the center to the left), and high-skilled workers move from low-productive firms to the high-productive firms (from the center to the right).

As it was stated before, the increase in wage inequality was mainly driven by the sharp increase of top wages. Figure 1 shows the top 10% and the top 1% wage shares in the U.S. from 1970 to 2010. The top 1% share increased from 5.1% in 1970 to 10.9% in 2010, and the top 10% share increased from 25.7% to 34.5%. The shares in 1970 were targeted in the calibration, but the shares in 2010 were not. The results of this experiment show how much of the overall increase in top wages can be explained by the additional channel in the labor market and an increase in the degree of complementarity in production. The results are presented in Table 4. In the model, the top 1% share increases by 3.41%, from 4.38% to 7.79%, while in the data it increases by 5.8%. The model is able to explain 59% of the actual increase in the top 1% wage share. For the top 10% wage share, the model predicts 8.58% increase, while the actual increase is 8.8%. The model accounts for 98% of the actual increase in the top 10% wage share.

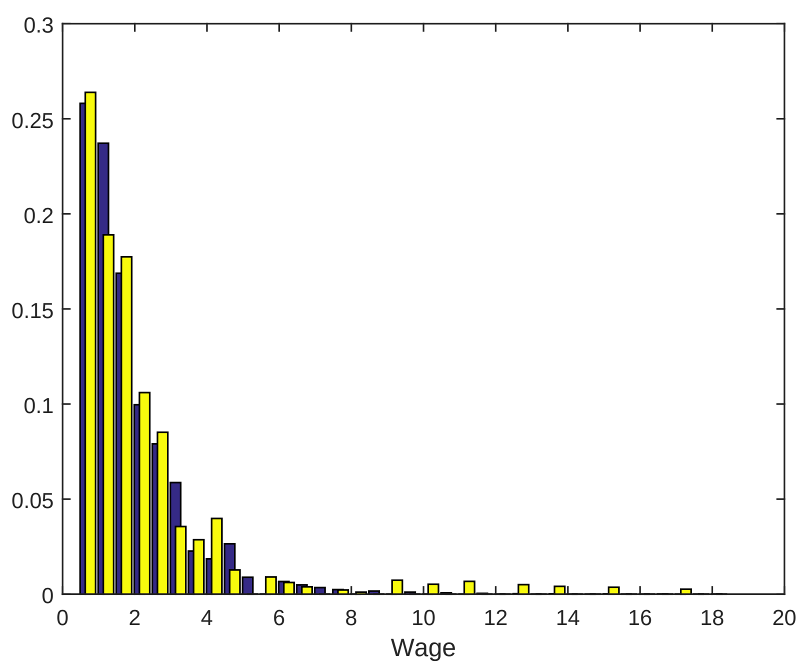
Figure 3: Distributions of wages.

(a) Distribution of wages without the headhunter channel.



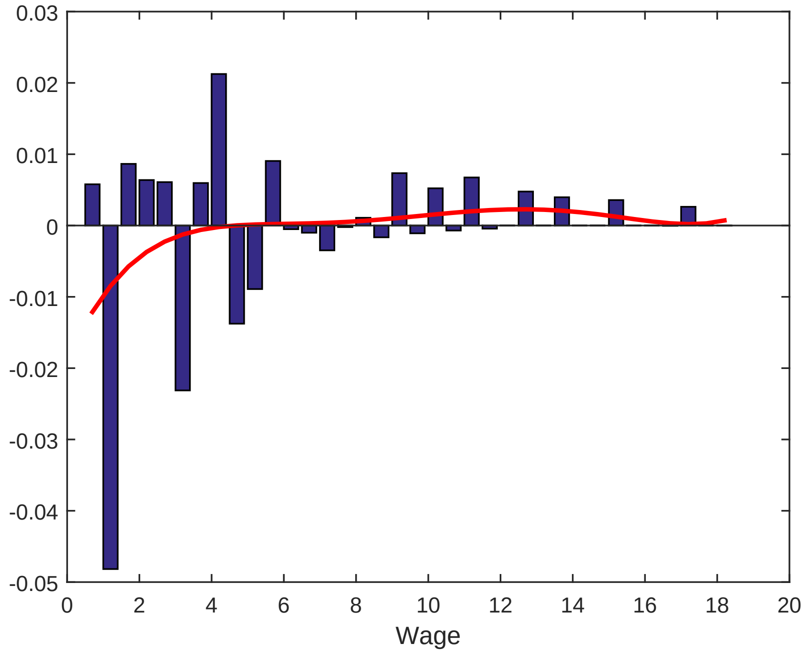
This figure plots the histogram of the distribution of wages in the model without the headhunter channel, baseline calibration.

(b) Distribution of wages with the headhunter channel.



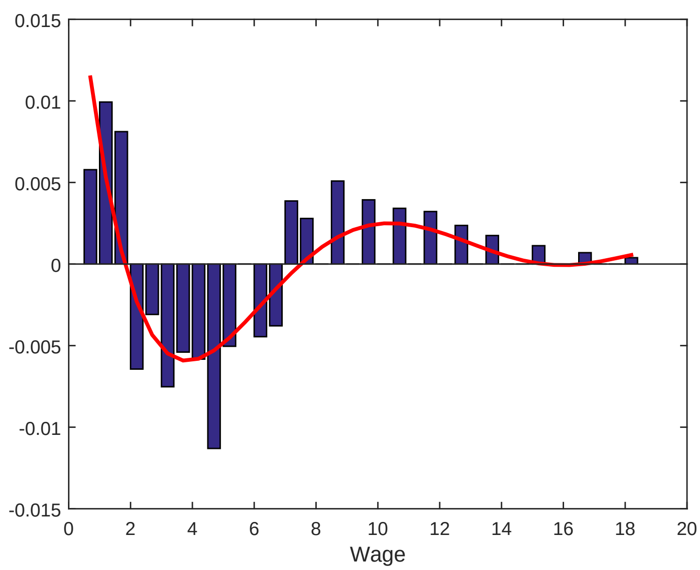
This figure plots the histogram of the distribution of wages in the model without the headhunter channel (blue bars) and the histogram of the distribution of wages in the model with the headhunter channel (yellow bars), baseline calibration.

(c) Difference between the distributions.



This figure plots the difference between the distributions of wages in the model with headhunters and without headhunters (blue bars). The red line plots a fitted polynomial of degree 5. Positive values indicate a larger mass of workers with a given wage in the model with the headhunter channel.

Figure 4: Difference between the distributions without SBTC.



This figure plots the difference between the distributions of wages in the model with headhunters and without headhunters (blue bars). The red line plots a fitted polynomial of degree 5. Positive values indicate a larger mass of workers with a given wage in the model with the headhunter channel.

Table 4: Top wage shares in the model and data.

Model	Top 1%	Top 10%	Data	Top 1%	Top 10 %
Without HH	4.38%	25.56%	1970	5.1%	25.7%
With HH	7.79%	34.14%	2010	10.9%	34.5%

This table presents the top wage shares in the modal and in the data. Columns 2 and 3 present the results from the model for the baseline calibration without the headhutner channel and the baseline calibration with the headhutner channel. Top 1% and 10% wage shares and 90/50 wage ratio in the U.S. in 1970 and 2010 are presented in columns 5 and 6 (source: Piketty (2014)).

4.3 Skill-Biased Technological Change

The large effect in Table 4 comes from skill-biased technological change and the headhunter channel acting together. To assess the relative contributions of the headhunter channel and the skill-biased technological change to the increase of the top wages I change separately only the matching technology (headhunter channel) or the degree of complementarity (SBTC). I present the results in the Table 5. First, I fix the degree of complementarity on the level of 1970 but add the headhunter channel (bottom-left panel). In this case the top 1% wage share is 6.65% ,instead of 7.79% in the baseline calibration (upper-left), and the top 10% wage share is 31.05% (instead of 34.14%). The headhunter channel contributes to explaining of 40% of the increase of top 1% wage share in the data and 70% of the increase in top 10% wage share. The headhunter channel also contributes to explain half of the rise of the upper-middle class, the 90/50 wage ratio rises by 0.20, while in the baseline calibration the rise is 0.39.

Table 5: Relative contribution of headhunters and SBTC.

<table>
<tr><th></th><th></th><th>HH</th><th>no HH</th></tr>
<tr><td rowspan="3">SBTC</td><td>Top 1%</td><td>7.79%</td><td>4.98%</td></tr>
<tr><td>Top 10%</td><td>34.14%</td><td>27.61%</td></tr>
<tr><td>Δ 90/50</td><td>0.39</td><td>0.16</td></tr>
<tr><td rowspan="3">no SBTC</td><td>Top 1%</td><td>6.65%</td><td>4.38%</td></tr>
<tr><td>Top 10%</td><td>31.05%</td><td>25.56%</td></tr>
<tr><td>Δ 90/50</td><td>0.20</td><td>0</td></tr>
</table>

This table presents the top wage shares and the change in 90/50 wage ratio for the baseline calibration with headhunters (upper-left), the calibration with SBTC but without headhunters (upper-right), the calibration without SBTC but with headhunters (lower-left), and the calibration without headhunters and no SBTC (lower-right).

If instead, I just increase the degree of complementarity to the level of the baseline calibration without the headhunter channel (upper-right), the top 1% wage share increases just to 4.98% and the top 10% wage share increases to 27.61%. The relative contribution of the degree of complementarity is about 10% out of 59% for top 1% wage share, and 23% out of 98% for the top 10% wage share. SBTC also explains almost another half of the increase of the upper-middle class.

Interaction between SBTC and the headhunter channel is also very important. The interaction explains around 9% of the increase in top 1% wage share (59%-40%-10%) and 5% of the increase in top 10% wage share (98%-70%-23%). With higher degree of complementarity, the importance of having a better match increases. Relative productivity of a firm with a high-skilled worker is even higher with respect to a similar firm with a low-skilled worker in case of high degree of complementarity. Better assortative matching on the top reinforces the effects of SBTC.

To give a chance to the SBTC to explain a higher proportion of the rise in top shares I recalibrate the SBTC to match the increase in the top 10% wage share or 90/50 wage ratio without the headhunter channel. I present the results in Table 6. We can see that the degree of complementarity must increase up to 1.24 without the headhunter channel to match the increase in 90/50 wage ratio, and up to 1.41 to match the top 10% wage share. When I match the 90/50 wage ratio, the model explains relatively small proportion of the increase in top 1% wage share (26%) and a significant proportion of the top 10% wage share (57%). If the increase in top 10% wage share is comparable to the baseline (or the headhunter channel alone), only SBTC loses a lot for the top 1% wage share. If I match the top 10% wage share, instead, the model explains a larger part of the top 10% wage share (all of it was targeted, while the headhunter channel alone explains 70%) and only slightly smaller increase in the top 1%. However, in this case the model predicts a very large increase in the 90/50 wage ratio (0.72) that is 82% higher than the actual increase. The reason for this is that the rise of the degree of complementarity alone rises all the wages uniformly and the rise must be enormous to match the top 10% wage share. We can see it in Figure 5. While with the headhunter channel all the wages rise only slightly due to higher degree of complementarity and the top wages rise more than that due to improvements in the assortative matching. With the headhunter channel, the high-skilled workers both move up with the curve and move along it to the right, and without the headhunter channel they can only move up with the curve.

These experiments show that the skill-biased technological change helps to explain the rise in the 90/50 wage ratio that corresponds to the rise of the upper-middle class relative to the bottom, but fails to replicate the sharp increase on the very top. The headhunter channel, instead, has the main effect on the top wages, rather than on the upper-middle class. Skill-biased technological change stretches the whole distribution to the right, while the headhunter channel fixes the left part of the distribution and moves the right tail further apart.

Table 6: Alternative calibration of SBTC.

	Data	Baseline	no HH, SBTC - 90/50	no HH, SBTC top 10
	2010	$\gamma = 1.10$	$\gamma = 1.24$	$\gamma = 1.41$
Top 1%	10.9%	7.79%	5.93%	7.26%
Top 10%	34.5%	34.14%	30.61%	34.43%
Δ 90/50	0.39	0.39	0.40	0.72

This table presents the top wage shares and the change in 90/50 wage ratio for the U.S. (column 2, source: Piketty (2014)) baseline calibration with headhunters (column 3), the calibration without headhunters and SBTC calibrated to match the change of 90/50 wage ratio (column 4), and the calibration without headhunters and SBTC calibrated to match the top 10

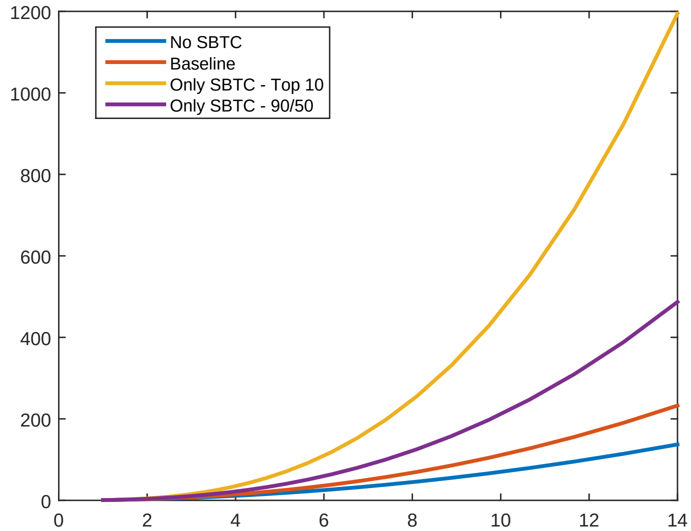
4.4 Assortative matching

The main mechanism behind the increase in wage inequality in the model is the increase in sorting between workers and firms, especially at the very top. With headhunters, high-skilled workers have exclusive opportunity to be matched with a high-productive firm, and high-productive firms, instead, have exclusive opportunity to be matched with high-skilled workers. Empirically, there are two most used ways to look at the assortative matching between workers and firms. First, one can directly compare the joint distributions of worker-firm matches over estimated types. And second, one can just look at the correlation between the types. I compute both statistics using the data simulated from the model in the baseline calibration in order to compare them to empirical estimates in the literature. The major drawback of this experiment, however, is that I can observe the real type of workers and firms directly, while in the data it is impossible.

First, I study the change in the joint distribution of worker-firm matches. To do it I split workers and firms into ten categories by their skill or productivity level and plot the joint distribution before and after introducing headhunters in the model. Figure 6a shows the distribution without headhunters, Figure 6b show the distribution with headhunters, and Figure 6c shows the change of the distribution. Numbers 1,2,3,..,10 in the figure correspond to the firm type, with 1 being the least productive firms and 10 being the most productive firms, and the color corresponding to the type of workers, with dark blue being the least skilled and yellow being the most skilled workers.

As it can be seen from the figures, with the baseline calibration of the model almost all high-skilled workers (within the top 10%) start working in the best firms (top 10%). All other firms lose significantly in the share of top workers and gain in the share of lower-skilled workers. This pattern is strikingly similar to the findings of Song et al. (2016) who plot similar distribu-

Figure 5: Wage of a worker in a best-fit firm.



This figure plots for each worker type (horizontal axis) her wage in the best-fit firm (vertical axis). Blue line represents calibration without SBTC, red line represents the baseline calibration, yellow line represents the calibration without the headhunter channel and the degree of complementarity changed in order to match the top 10% wage share in 2010, and purple line represents the calibration without the headhunter channel and the degree of complementarity changed in order to match the change in 90/50 wage ratio between from 1970 to 2010.

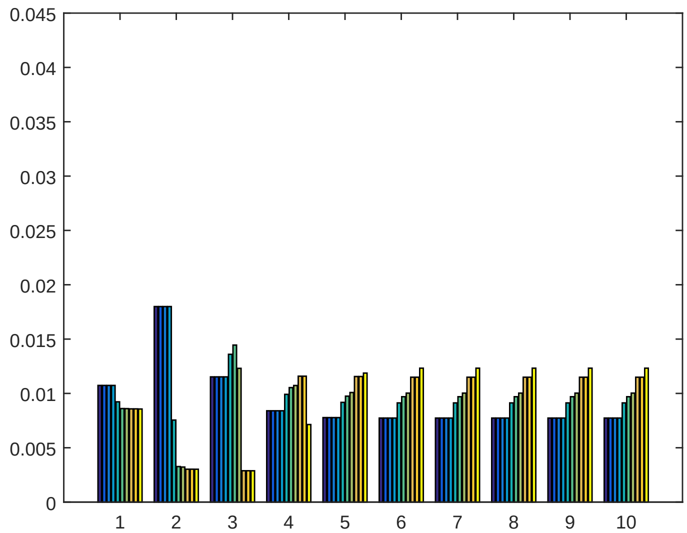
tions for estimated workers and firms fixed effects from the U.S. data. Comparing to the data, the model with the baseline calibration exaggerates the increase in the number of top workers working in top firms because most of the firms on top use the headhunter channel and the headhunters don't do mistakes screening workers. In the empirical counterpart, there is an error in estimating the true type of the worker that should smooth the figure, and headhunters might do mistakes by assessing the skill of workers with mistakes, therefore smoothing even more the resulting difference.

The second way to analyze sorting in the labor market is to compute correlations between the types of the workers and firms. In order to do it in the model simulated data, I draw 100000 matches from the joint worker-firm distribution in the model and decompose the variance of log wages into worker type, firm type, and covariance between the two. Table 7 presents the results of this experiment for the steady-state without headhunters, with headhunters, and the difference between the two.

From the results of this experiment ,we can see that covariance and correlation of the worker and firm types increase significantly after the introduction of headhunters to the model. Indeed, the increase is not only in the top part of the distribution but over the whole distribution. This

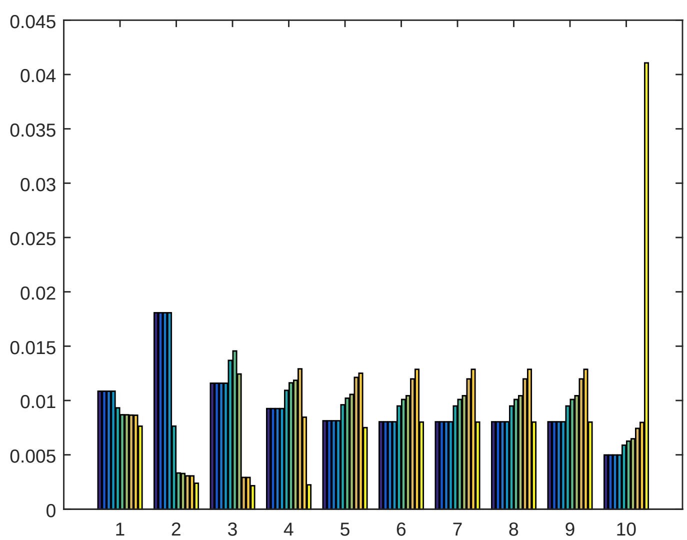
Figure 6: Joint distributions of worker-firm matches.

(a) Joint distribution without the headhunter channel.



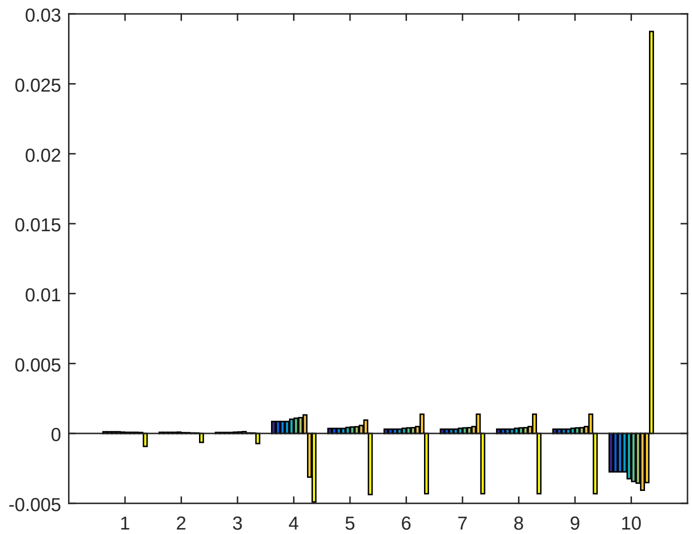
This figure plots the joint distribution of the matches of workers and firms over deciles of workers' and forms' types in the baseline calibration without the headhunter channel. Numbers 1-10 represent the firm type deciles with 1 being the least productive and 10 the most productive firms. Colors represent the worker type deciles with dark blue being the least skilled workers and bright yellow the most skilled workers.

(b) Joint distribution with the headhunter channel.



This figure plots the joint distribution of the matches of workers and firms over deciles of workers' and forms' types in the baseline calibration with the headhunter channel. Numbers 1-10 represent the firm type deciles with 1 being the least productive and 10 the most productive firms. Colors represent the worker type deciles with dark blue being the least skilled workers and bright yellow the most skilled workers.

(c) Change in the joint distribution.



This figure plots the difference between the joint distributions of the matches of workers and firms over deciles of workers' and forms' types in the baseline calibration with and without the headhunter channel. Positive values indicate larger number of matches in the calibration with the headhunter channel. Numbers 1-10 represent the firm type deciles with 1 being the least productive and 10 the most productive firms. Colors represent the worker type deciles with dark blue being the least skilled workers and bright yellow the most skilled workers.

happens because the high-skilled workers that move to the best firms free the positions for low- and medium-skilled workers in the rest of the firms, also improving the matching for them. Again, this result is in line with the empirical results by Song et al. (2016) who show that increased covariance between worker and firm types was one of the major drivers in the increase in the variance of wages in the U.S. between the 1980s and 2000s.

Improvement in assortative matching naturally translates in higher aggregate production in the economy (Table 8). Better matches at the top significantly improve average productivity in the economy. In the baseline calibration this improvement is as much as 22% together with the increase because of higher degree of complementarity. To assess the contribution of the matching alone, we can look at the distance of the allocation to the first best. In the calibration without the headhunter channel, the aggregate production is 88% of the first best allocation of workers to firms, while in the calibration with headhunters it is already 91.5%. The improvement in the matching brings the economy by 3.5% closer to the first best allocation, that is a 5-7% increase in production.

Table 7: Log wage variance decomposition and correlation of worker and firm types.

	Without HH	With HH	Dif	%
$Var(\log(w))$	0.3020	0.4262	0.1242	100
$Var(\log(e))$	0.1289	0.1549	0.0260	21%
$Var(\log(p))$	0.1302	0.1567	0.0265	21%
$2Cov(\log(e), \log(p))$	0.0428	0.1147	0.0719	58%
$Cor(\log(e), \log(p))$	0.1656	0.3679	0.2023	-

This table presents the results of the wage variance decomposition based on the 100000 draws from the model-simulated data with and without the headhunter channel.

Table 8: Aggregate production.

	no HH	of FB	HH	of FB	HH-noHH	%	without SBTC	FB
no SBTC	2.82	88.13%	2.97	92.81%	0.15	5.3%	-	3.20
SBTC	3.21	85.6%	3.43	91.47%	0.22	6.85%	0.61/21.63%	3.75

This table presents the aggregate production for the calibration of the model without and with the headhunter channel.

4.5 Robustness

To check how much the magnitude of the increase depends on the choice of the skill threshold for the headhunter channel, I do similar experiments for top 1%, 2%, 3%, 7%, or 10% of workers being eligible to the headhunter channel. The results are presented in Table 9. Not surprisingly, higher thresholds, except for top 1% threshold, (more efficient screening by headhunters) increase the wage share of top 1% of the workers but decrease the share of top 10%. This happens because, with a higher threshold, the most efficient workers are more concentrated in the top firms, for example, they all work in top 2% of the firms instead of top 5%. Their wages increase even more due to complementarities, so the top 1% wage share increases more. Instead, for the workers in the 10-1% bracket, the probability of working in the best firms decreases with a higher threshold. Workers in 5-2% are excluded from the headhunter channel and many of them end up in bad or average firms, so the top 10% wage share drops relative to baseline calibration despite the top 1% wage share increase. The overall effect of the sorting mechanism is still striking - it explains at least 40% of the actual increase in the top 1% share of wages and 55%% of the top 10% wage share (together with SBTC). We can also see that with the threshold set to top 1% the increase in the top 1% wage share drops relative to the top 2% threshold. This happens because not all firms in top 1% use the headhunter channel and therefore only around a half of the workers in the top 1% are matched through the headhunter channel.

Another target of choice that doesn't have a properly estimated empirical counterpart is the

Table 9: Top wage shares in the model and data for other skill thresholds.

Model	Share	Top 1%	Top 10%	Δ 90/50
Without HH	0	4.38%	25.56%	0
With HH on top 5% (basline)	5.05%	7.79%	34.14%	0.39
With HH on top 10%	10.62%	6.70%	35.81%	0.65
With HH on top 7%	7.47%	7.00%	35.08%	0.39
With HH on top 3%	3.27%	8.01%	32.99%	0.39
With HH on top 2%	2.00%	9.15%	31.51%	0.39
With HH on top 1%	1.16%	8.44%	30.42%	0.16

This table presents the results of the main experiment with alternative calibrations of the skill threshold. The results are presented for the skill thresholds of 5% (baseline calibration), 3%, 2%, and 1%.

share of firms using the headhunters. In the baseline calibration, it is chosen to fit the estimates by AESC, but I also redo the experiment with different shares to see how sensitive is the result depending on the choice of the target. I set the share of the firms using the headhunter channel to be 20%, 40%, 60%, 80%, or 100%. The results are presented in Table 10. Of course, the increase of the top wage shares is decreasing with a lower share of firms using headhunters. But the major part of the effect is still there even if every 5th firm is allowed to use the headhunter channel every period. In this case, the model is still able to explain 53% of the increase in top 1% wage share and 80% of the increase in the top 10% wage share (again, together with SBTC). Even when the share of firms using headhunters is set to the most conservative estimate the model is still able to predict a large share of the increase in top wages.

Table 10: Top wage shares in the model and data for different intensity of the use of headhunter channel.

Model	Top 1%	Top 10%	Δ 90/50
Without HH	4.38%	25.56%	0
With HH, baseline	7.79%	34.14%	0.39
With HH, 100%	9.00%	36.22%	0.39
With HH, 80%	8.12%	34.77%	0.39
With HH, 60%	7.90%	34.28%	0.39
With HH, 40%	7.56%	33.80%	0.39
With HH, 20%	7.46%	32.65%	0.39

This table presents the results of the main experiment with alternative calibrations of the distributions of non-monetary costs of headhunters. The results are presented for the share of firms using headhunters of 54% (baseline), 100%, 80%, 60%, 40%, and 20%.

4.6 Discussion

As it was discussed before, the main effect of the headhunter channel on the wage distribution comes from the separation of the labor markets for high-paying and low-paying jobs. With the headhunter channel in place, only high-skilled workers get to high-paying jobs and their wages increase dramatically due to skill and firm productivity complementarities. Moreover, headhunter channel allows high-skilled workers to search on-the-job less costly because they don't have to pay the search cost every period but only when they receive a call from the headhunter. Because of this, high-skilled workers agree to consider an offer even when they work in medium- and high-productive firms, while they would stop searching for a job actively working in such firms without the headhunter channel. Low-skilled workers, in contrast, lose their possibility to work in high-paying jobs, therefore their wages are compressed to lower levels after the introduction of the headhunters.

Simulations presented in this section show that the change of the matching technology can explain a major part of the increase in the wage inequality. The results are quite robust to different calibrations of the headhunter channel and are amplified by skill-biased technological change in the model. Skill-biased technological change, instead, has its main effect on the upper-middle class, contributing significantly to the rise of the 90/50 wage ratio and having a smaller contribution to the rise of top wages. The main reason for this is that the headhunter channel creates the non-linearity in the matching pattern between workers and firms, as observed in the data. SBTC alone gradually improves sorting across all distribution without the sharp increase of top workers in top firms. Only this sharp increase due to strong non-linearity rises the top wages much higher than the middle and low wages.

5 Cross-country comparison

This section discusses the cross-country evidence supporting the main mechanism of the paper, that the headhunters by improving matching on the labor market increase the top wages. Headhunters entered labor market of different countries in different periods and therefore are used by firms to a different extent[16]. This variation in headhunters' activity helps to establish causality between the role of headhunters and the growth of top wages. This section uses the

[16]Different labor market legislation is another source of exogenous differences of headhunter activity across countries.

data on the major headhunter companies in European labor markets in 1997 and the data on top income shares between 1980 and 2010. Ideally, top wage shares should be used in the analysis, however, such data is not available for all countries over the whole period in question.

Data on major headhunter companies operating in Europe is available in Jenn (1999). The data includes the distribution of fee revenues as well as the number of hires by country[17]. I aggregate the data by country to get total fees and a total number of hires by headhunters in a country in 1997, and I normalize the data by GDP in 1997 (for fee revenues) or total employment in 1997 (for hires). Normalization allows comparing the share of headhunters between countries. The question that this analysis answers is what is the relation between headhunter activity and the dynamics of top incomes? To answer this question figures 7 and 8 plot the relations between normalized hires by headhunters and top income shares, or growth of top income shares. Figures 9 and 10 plot similar relations for normalized fee revenues. Figures 7a and 8a show that there is a strong positive correlation between normalized hires by headhunters and the future growth of top incomes. Only Norway falls from the general pattern, but Norway experienced a change in capital income taxation in 2006, so most likely this drop is not related to labor incomes.

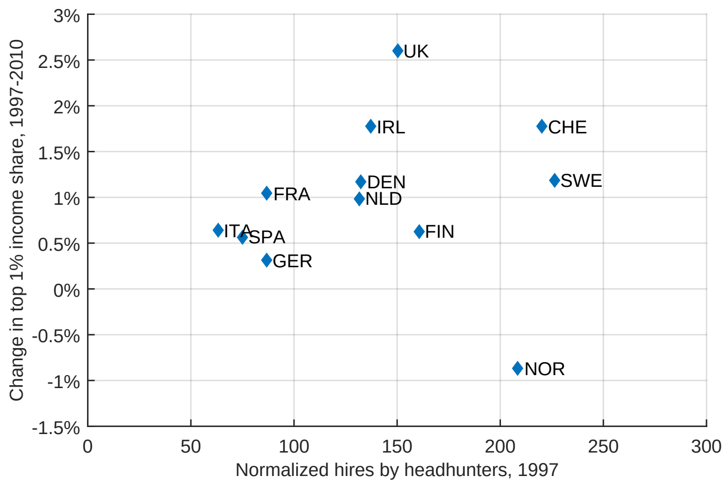
To address the concern that headhunters were more active in countries where top wages were already higher, figures 7b and 8b plot the relation between top income shares in 1997 and normalized hires by headhunters in 1997. As it is evident from the figures, there is no correlation between headhunter activity and top income shares in 1997. It means that differences of headhunter intensity across countries are driven by other factors, exogenous to top income levels. To further strengthen this claim, figures 7c and 8c plot the top income shares growth before 1997 against normalized hires in 1997. Lack of positive correlation shows that headhunters intensity is not driven by the previous growth of top incomes. Headhunters didn't choose countries with fast growing top incomes.

Figures 9 and 10 present similar analysis for normalized fee revenues. All the results for normalized hires hold also for normalized fee revenues. This analysis shows two important facts. First, it shows that headhunters indeed signal future growth of top incomes. In the model the increase of top incomes happens because of improved matching at the top, with headhunters inducing the better matching. This evidence, however, doesn't provide any hints on the mechanism of the top incomes increase, or the degree of quality of the matching. Second, this analysis shows that the distribution of headhunter activity over countries is exogenous to the level of

[17] For some companies the number of hires is estimated and not exactly observed.

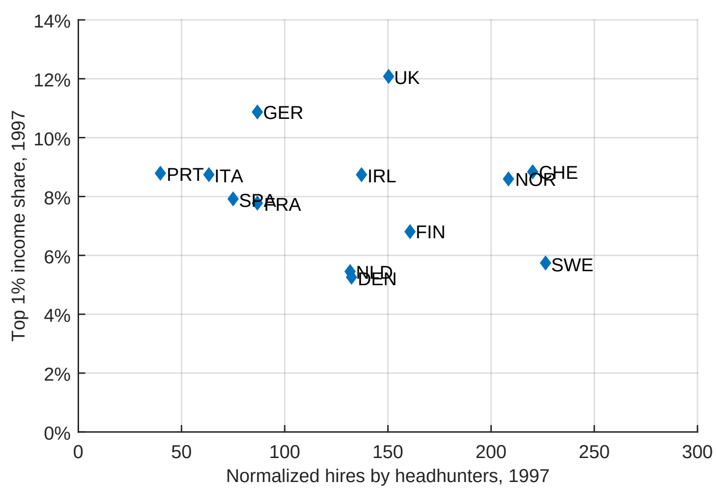
Figure 7: Top 1% income share and normalized hires by headhunters.

(a) Growth of top 1% income share, 1997-2010, and normalized hires by headhunters, 1997.



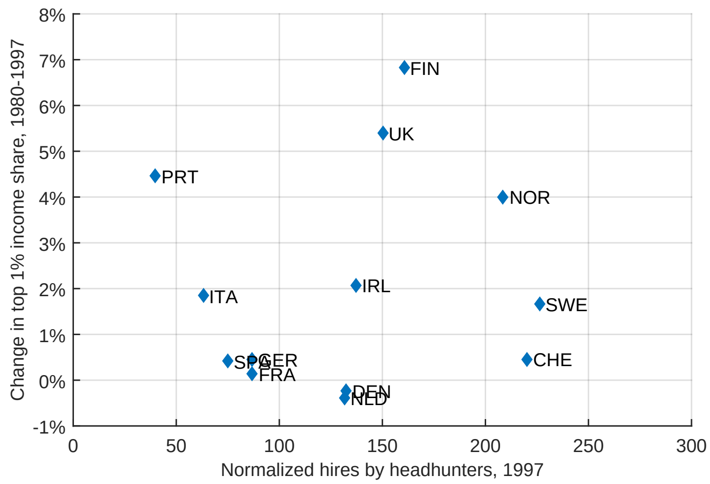
This figure plots the change in the top 1% income share in a country between 1997 and 2010 (vertical axis) against the hires by headhunters in 1997 normalized by total employment in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

(b) Top 1% income share, 1997, and normalized hires by headhunters, 1997.



This figure plots the top 1% income share in a country in 1997 (vertical axis) against the hires by headhunters in 1997 normalized by total employment in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

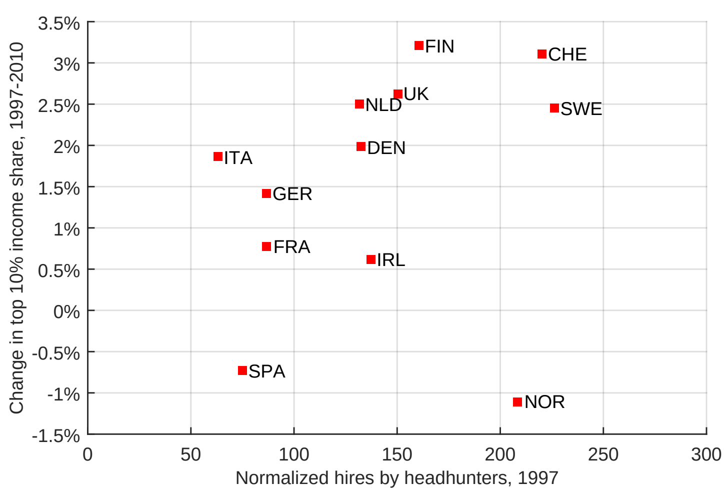
(c) Growth of top 1% income share, 1980-1997, and normalized hires by headhunters, 1997.



This figure plots the change in the top 1% income share in a country between 1980 and 1997 (vertical axis) against the hires by headhunters in 1997 normalized by total employment in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

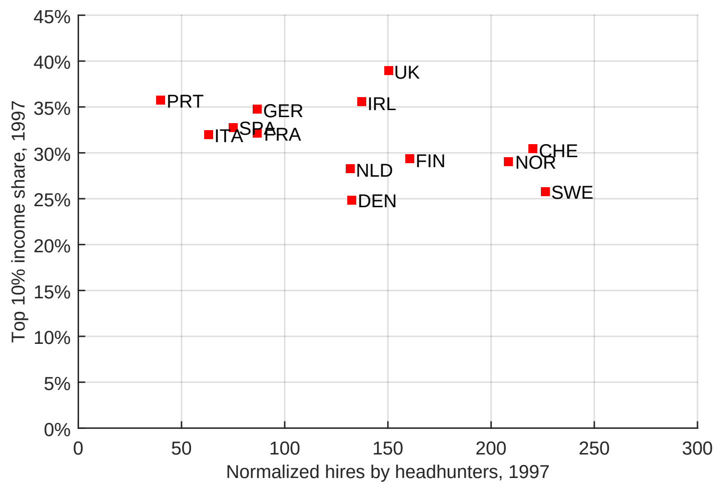
Figure 8: Top 10% income share and normalized hires by headhunters.

(a) Growth of top 10% income share, 1997-2010, and normalized hires by headhunters, 1997.



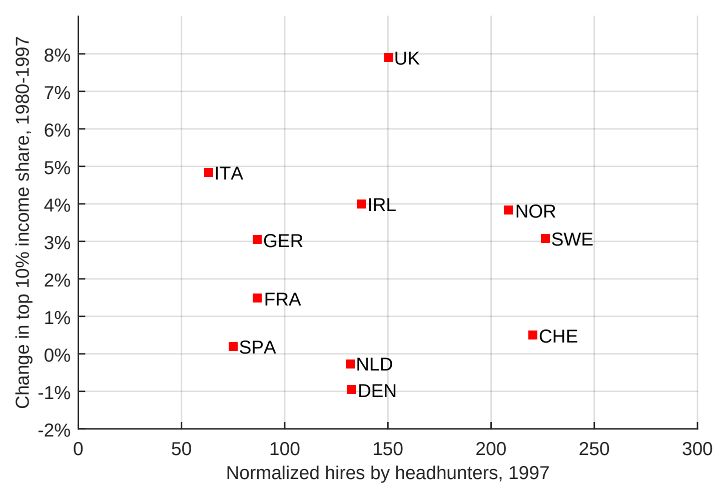
This figure plots the change in the top 10% income share in a country between 1997 and 2010 (vertical axis) against the hires by headhunters in 1997 normalized by total employment in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

(b) Top 10% income share, 1997, and normalized hires by headhunters, 1997.



This figure plots the top 10% income share in a country in 1997 (vertical axis) against the hires by headhunters in 1997 normalized by total employment in 1997(horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

(c) Growth of top 10% income share, 1980-1997, and normalized hires by headhunters, 1997.



This figure plots the change in the top 10% income share in a country between 1980 and 1997 (vertical axis) against the hires by headhunters in 1997 normalized by total employment in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

top incomes and the history of the growth of top incomes. There must be other factors limiting headhunter activity in some countries, for example, labor market legislation, or high costs of establishing detailed databases of potential candidates.

The importance of this empirical evidence is in demonstrating the lack of reverse causality. Headhunters might be more active in countries where the income inequality was higher so they came to the market to extract higher fee revenues. In this case, the increasing top wages would drive the rise in the headhunter industry, and the mechanism presented in this paper would not be present. However, the results presented in this section show that only the future change in top incomes is correlated with the headhunter intensity, so reverse causality can be rejected.

6 Micro evidence

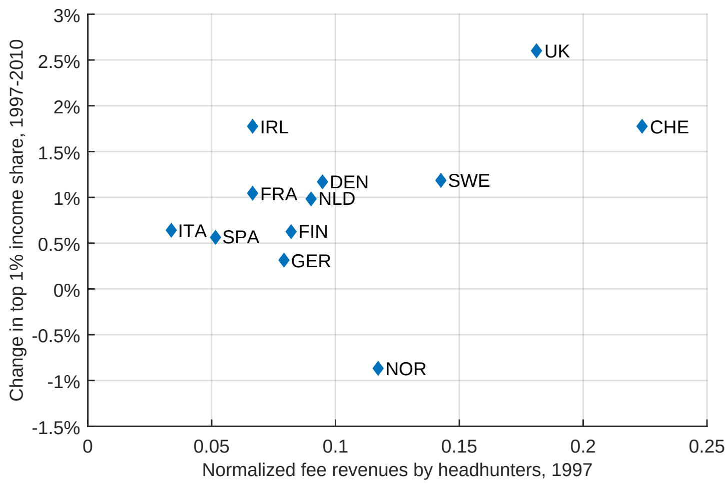
This section presents the empirical analysis of the potential effect of headhunters on the CEO compensation and firm performance. CEOs constitute a major part of the hires by headhunters, accounting for 20000 hires by headhunters in 2013 in the U.S. alone, and therefore are a good proxy for individual effects of headhunters on matching between workers and firms. There is no information on the identity of the CEOs that were hired through headhunters, so the evidence presented here should not be considered as a direct proof of the work of the mechanism. All conclusions are suggestive.

6.1 Data

The data on CEO characteristics and compensation and the firm level data are obtained from COMPUSTAT dataset. Following Gabaix, Landier, and Sauvagnat (2014), four proxies for the firm size will be used, constructed from variables obtained from COMPUSTAT yearly data set. First, firm value is constructed as the sum of the market value of equity, defined as a number of shares outstanding multiplied by the end-of-fiscal-year stock price, and the book value of debt, defined as total assets minus the sum of the book value of equity and deferred taxes. Second, equity value constructed as the number of shares outstanding multiplied by the end-of-fiscal-year stock price. Third, the sales variable from the COMPUSTAT. Fourth, the income is measured as earnings before interest and taxes. For CEO compensation EXECUCOMP panel of the COMPUSTAT will be used that contains information about the top 5 paid executives

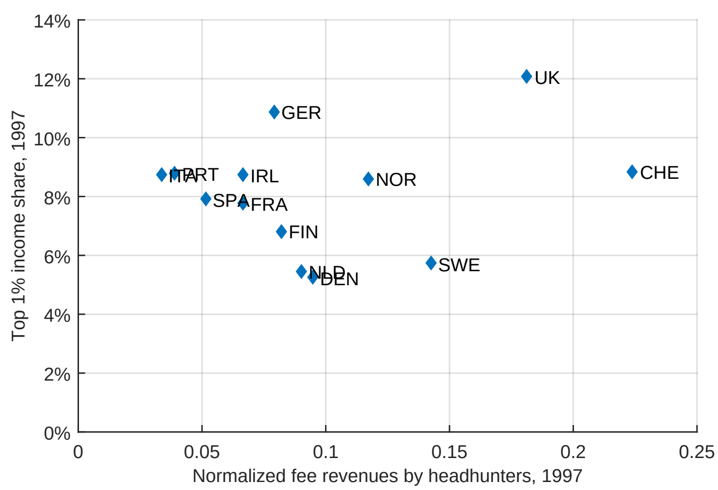
Figure 9: Top 1% income share and normalized fee revenues by headhunters.

(a) Growth of top 1% income share, 1997-2010, and normalized fee revenues by headhunters, 1997.



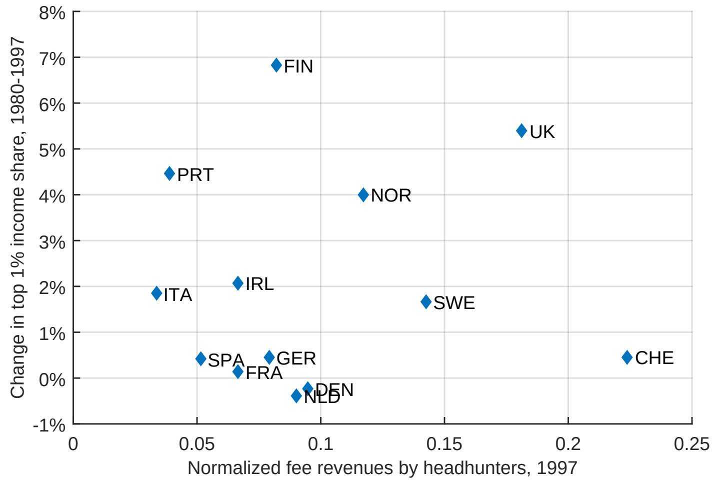
This figure plots the change in the top 1% income share in a country between 1997 and 2010 (vertical axis) against the fee revenues by headhunters in 1997 normalized by GDP in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

(b) Top 1% income share, 1997, and normalized fee revenues by headhunters, 1997.



This figure plots the top 1% income share in a country in 1997 (vertical axis) against the fee revenues by headhunters in 1997 normalized by GDP in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

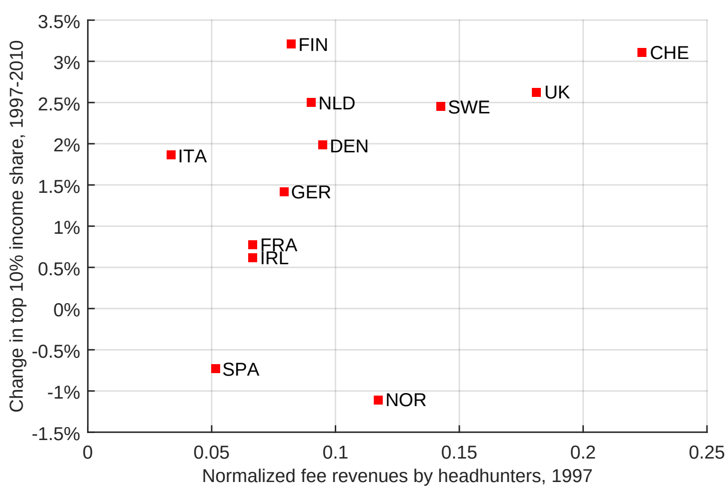
(c) Growth of top 1% income share, 1980-1997, and normalized fee revenues by headhunters, 1997.



This figure plots the change in the top 1% income share in a country between 1980 and 1997 (vertical axis) against the fee revenues by headhunters in 1997 normalized by GDP in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

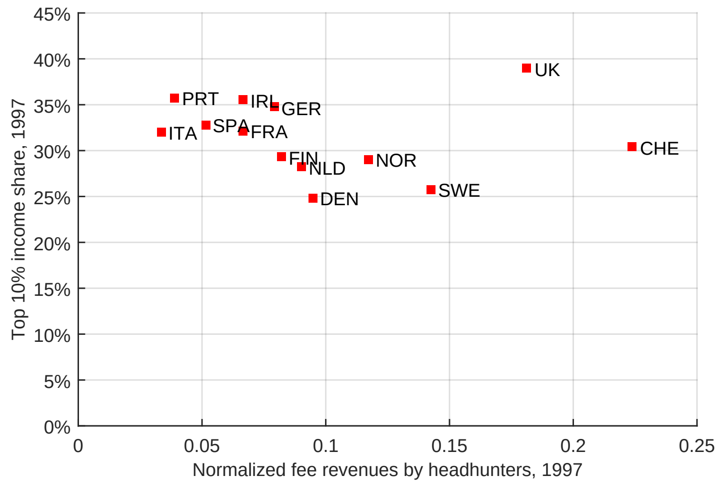
Figure 10: Top 10% income share and normalized fee revenues by headhunters.

(a) Growth of top 10% income share, 1997-2010, and normalized fee revenues by headhunters, 1997.



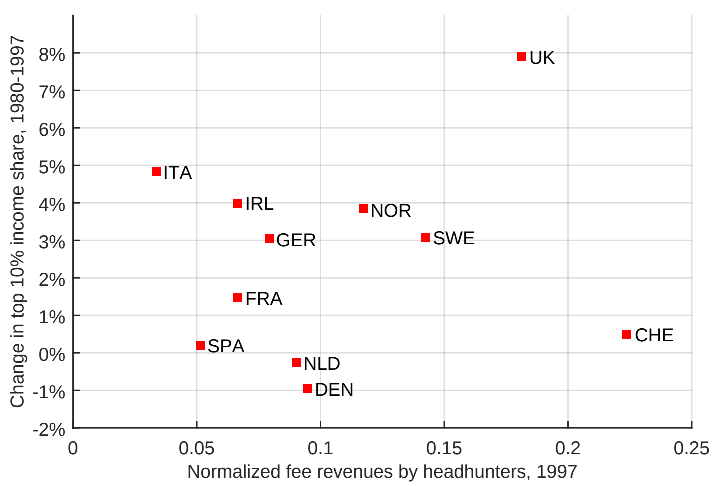
This figure plots the change in the top 10% income share in a country between 1997 and 2010 (vertical axis) against the fee revenues by headhunters in 1997 normalized by GDP in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

(b) Top 10% income share, 1997, and normalized fee revenues by headhunters, 1997.



This figure plots the top 10% income share in a country in 1997 (vertical axis) against the fee revenues by headhunters in 1997 normalized by GDP in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

(c) Growth of top 10% income share, 1980-1997, and normalized fee revenues by headhunters, 1997.



This figure plots the change in the top 10% income share in a country between 1980 and 1997 (vertical axis) against the fee revenues by headhunters in 1997 normalized by GDP in 1997 (horizontal axis). Sources: http://www.wid.world/, Jenn (1999), https://data.oecd.org/, and author's calculations.

of the largest firms in the U.S. In particular, the variable TDC1 will be used to measure CEO compensation, that includes salary, bonus, restricted stock granted and the Black-Scholes value of stock options granted. To construct industry dummies the four-digit SIC industry codes are used as in Fama and French (1997). A dummy variable for a change of CEO is constructed such that it is equal to 0 if the CEO is the same as the CEO of the first observation of the company, and 1 otherwise:

$$NewCEO_{i,t} = \begin{cases} 1 & \text{if CEO is different from the first observation of the firm} \\ 0 & \text{otherwise.} \end{cases}$$

Another important variable that will be taking into account is the index of enforceability of non-competition constructed by Garmaise (2009). The index is higher in the states where the non-compete agreements are enforced by courts and low in the states were the non-compete agreements are forbidden. Non-compete agreements restrict job-to-job transitions for workers and therefore limit the activity of headhunters.

The following sample will be used. The maximum available time period in the data set is from 1993 to 2013. The analysis will be restricted only to the CEO of every company. If a firm changes the CEO more than ones in the sample period, all observations starting from the third CEO are dropped. These restrictions leave 3102 firms with 7.95 average years of observation. The reason for leaving only CEOs is motivated by two facts. First, CEO is the highest paid executive in almost all the firms in the sample, therefore they are more likely to be hired through a head-hunter. And second, it is perceived that CEOs have the most influence on the firm activities and performance, so matching with a good CEO improves the firm's performance.

6.2 Results

I estimate the following equation:

$$\log(TDC1_{i,t}) = \alpha * NewCEO_{i,t} + \beta * \log(Firm\ size_{i,t}) + FE_t + FE_i + \varepsilon_{i,t},$$

where $TDC1_{i,t}$ is CEO compensation in firm i and year t, $NewCEO_{i,t}$ is the dummy variable constructed as described above, and $Firm\ size_{i,t}$ is one of the four measures of the firm size described above in firm i and year t. Table 11 presents the results on the full sample described before. Columns (1) and (2) present the results of the estimation with firm fixed effects and with or without the year fixed effects. Columns (3) and (4) present the results for similar estimation but without the $NewCEO_{i,t}$ dummy variable. The obtained results show that after a company changes the CEO it pays her from 5% to 16% more than to the previous CEO after conditioning on the firm size.

Table 11: CEO compensation and change of the CEO, full sample.

	Sample period 1993-2013			
Log of compensation	(1)	(2)	(3)	(4)
New CEO	0.1579***	0.0495**	-	-
	(0.0296)	(0.0180)	-	-
Log of Firm Value	0.1449***	0.0896**	0.1700***	0.0883**
	(0.0395)	(0.0391)	(0.0410)	(0.0393)
Log of Equity Value	0.1936***	0.2243***	0.1763***	0.2234***
	(0.0286)	(0.0310)	(0.0291)	(0.0310)
Log of Income	0.0852***	0.0899***	0.0761***	0.0898***
	(0.0135)	(0.0118)	(0.0138)	(0.0119)
Log of Sales	0.0427	0.0122	0.0809***	0.0105
	(0.0253)	(0.0259)	(0.0809)	(0.0259)
Year FE	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes
R^2	0.66	0.66	0.66	0.66
Number of observations	24673	24673	24673	24673

* p<10%, ** p<5%, *** p<1%

.

More interestingly, if we focus on two sub-periods in the sample (Table 12) - 1993 to 1998 and 2004 to 2007 - the periods during which the headhunter revenues were increasing particularly fast, as it can be clearly seen from the Figure 2a, the coefficient of interest increases from 5% to 9% in the first sub-period and from 5% to 13,6% in the second sub-period. This can be viewed as an indirect evidence of the higher use of headhunters during that periods and therefore improvements in the matching between CEOs and firms that resulted in higher compensation.

Table 13 presents the results for individual measures of the firm size. Columns (1) and (2) present the results for the firm value measure and columns (3) and (4) for the equity value measures. The results are consistent with the results for the full sample.

The important question is what is the channel of this effect, why are the firms paying more to the new CEOs. One potential explanation can be that the new CEO has a higher bargaining power than the previous CEO, it can be the case especially if the new CEO was hired with the help of a headhunter while the previous CEO was not. To test for this channel I augment the estimated equation with the interaction term between the $NewCEO$ dummy variable and the

Table 12: CEO compensation and change of the CEO, sub periods of headhunters revenue booms.

	1993-1998		2004-2007	
Log of compensation	(1)	(2)	(3)	(4)
New CEO	0.0906**	-	0.1364**	-
	(0.0315)	-	(0.0382)	-
Log of Firm Value	0.1648	0.1615	0.0497	0.0450
	(0.0943)	(0.0942)	(0.1662)	(0.1653)
Log of Equity Value	0.2359**	0.2370**	0.2487***	0.2521***
	(0.0713)	(0.0714)	(0.0415)	(0.0397)
Log of Income	0.1034***	0.1037***	0.1396**	0.1405**
	(0.0187)	(0.0190)	(0.0263)	(0.0271)
Log of Sales	-0.0542	-0.0665	-0.1549	-0.1697
	(0.0572)	(0.0567)	(0.0989)	(0.1024)
Year FE	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
R^2	0.724	0.723	0.785	0.785
Number of observations	8304	8304	4505	4505

* p<10%, ** p<5%, *** p<1%

.

measures of the firm size. The coefficient of the measure of the firm size in this regression can be viewed as the bargaining power of the CEO, i.e. how much his compensation increases when the firm is growing, and the interaction term can be viewed as the change in the bargaining power of the new and the previous CEOs. The results are presented in Table 14. Columns (1) and (2) present the results for the firm value as a proxy for the firm size, while columns (3) and (4) present the results for the equity value of the firm. The results show that the interaction term is negative or statistically not significantly different from 0. This shows that the increase of CEO compensation does not come from the higher bargaining power.

Now, to test for the channel that works in the model, the increased productivity of a match resulting in a higher wage, the direct effect of a change of CEO on the firm size is tested. Table 15 presents the results, with columns (1) and (2) showing the effect on the firm value and columns (3) and (4) showing the effect on the equity value. As it can be seen from the table the effect of the change on the CEO on the firm size is positive. This can be an indicator of the presence of the channel related to the productivity of the match between the new CEO and the firm.

To further explore the matching channel I add the non-competition enforceability index to the analysis. First, Table 16 presents the result of the regression including just the enforceability index, but not the new CEO dummy. It confirms the well-known result that CEO compensation

Table 13: CEO compensation and change of the CEO, individual firm size measures.

	Sample period 1993-2013			
Log of compensation	(1)	(2)	(3)	(4)
New CEO	0.0437**	0.1389***	0.0420**	0.1876***
	(0.0177)	(0.0297)	(0.0171)	(0.0314)
Log of Firm Value	0.4311***	0.4620***	-	-
	(0.0180)	(0.0193)	-	-
Log of Equity Value	-	-	0.3442***	0.3572***
	-	-	(0.0161)	(0.0191)
Year FE	Yes	No	Yes	No
Firm FE	Yes	Yes	Yes	Yes
R^2	0.66	0.653	0.66	0.654
Number of observations	24673	24673	24673	24673

* p<10%, ** p<5%, *** p<1%

.

Table 14: CEO compensation and change of the CEO, bargaining power.

	Sample period 1993 - 2013			
Log of compensation	(1)	(2)	(3)	(4)
New CEO	0.3590***	0.3145**	0.4079***	0.3168**
	(0.1293)	(0.1277)	(0.1318)	(0.1302)
Log of Firm Value	0.4746***	0.4466***	-	-
	(0.0173)	(0.0183)	-	-
Log of Equity Value	-	-	0.3720***	0.3629***
	-	-	(0.0159)	(0.0164)
Log of FV*New CEO	-0.0269	-0.0332**	-	-
	(0.0165)	(0.0165)	-	-
Log of EV*New CEO	-	-	-0.0291	-0.0366**
	-	-	(0.0182)	(0.0184)
Year FE	No	Yes	No	Yes
Firm FE	Yes	Yes	Yes	Yes
R^2	0.654	0.659	0.654	0.662
Number of observations	24673	24673	24673	24673

* p<10%, ** p<5%, *** p<1%

.

Table 15: CEO compensation and change of the CEO, match efficiency.

	Firm Value		Equity Value	
	(1)	(2)	(4)	(5)
New CEO	0.4784***	0.6138***	0.4023***	0.5208***
	(0.0399)	(0.0450)	(0.0399)	(0.0448)
Year FE	No	Yes	No	Yes
Industry FE	Yes	Yes	Yes	Yes
R^2	0.200	0.231	0.095	0.132
Number of observations	24673	24673	24673	24673

* p<10%, ** p<5%, *** p<1%

.

Table 16: CEO compensation and the non-compete enforceability index.

	Sample period 1993 - 2013		
Log of compensation	(1)	(2)	(3)
NCEI	-0.0222***	-0.0286***	-0.0221***
	(0.0028)	(0.0029)	(0.0029)
Log of Firm Value	0.1780***	0.0514***	0.1608***
	(0.0149)	(0.0146)	(0.0149)
Log of Equity Value	0.0982***	0.2114***	0.1025***
	(0.0184)	(0.0149)	(0.0183)
Log of Income	0.0640***	0.04322**	0.0691***
	(0.0122)	(0.0152)	(0.0114)
Log of Sales	0.1268***	0.1571***	0.1397***
	(0.0108)	(0.0082)	(0.0086)
Year FE	No	Yes	Yes
Industry FE	Yes	No	Yes
R^2	0.410	0.384	0.420
Number of observations	24217	24217	24217

* p<10%, ** p<5%, *** p<1%

.

is lower in the states that enforce the non-compete agreements.

Table 17 presents the result of the interaction between the non-compete enforceability index and the new CEO dummy. The interaction term increases the magnitude of the effect of the new CEO dummy and has a negative significant coefficient. It means that the effect of the change of the CEO on the compensation that the firm pays is higher in the states with low non-compete enforceability index. And it decreases with the higher index. It is also interesting to notice that in the states where the index would be 1 (the highest index is 0.9 in Florida) the overall effect of the change of CEO on the compensation would be negative.

Table 17: CEO compensation, new CEOs, and the non-compete enforceability index.

	Sample period 1993 - 2013		
Log of compensation	(1)	(2)	(3)
NCEI*New CEO	-0.1685***	-0.2377***	-0.1805***
	(0.0475)	(0.0477)	(0.0471)
New CEO	0.1245***	0.0778***	0.0663***
	(0.0282)	(0.0190)	(0.0191)
Log of Firm Value	0.1769***	0.0531***	0.1633***
	(0.0161)	(0.0151)	(0.0156)
Log of Equity Value	0.1031***	0.2184***	0.1049***
	(0.0180)	(0.0151)	(0.0183)
Log of Income	0.0626***	0.0390**	0.0673***
	(0.0121)	(0.0152)	(0.0114)
Log of Sales	0.1208***	0.1528***	0.1368***
	(0.0113)	(0.0086)	(0.0089)
Year FE	No	Yes	Yes
Industry FE	Yes	No	Yes
R^2	0.409	0.381	0.419
Number of observations	24217	24217	24217

* p<10%, ** p<5%, *** p<1%

.

6.3 Discussion

The empirical results presented in this section show that the matches between CEOs and companies improve over time in the U.S. The improvement in the matching results in the higher CEO compensation and larger size of the firms. This is important in the context of this paper because the CEOs constitute a major part of the hires by headhunters, and because CEOs are exactly the people on the top of the wage distribution. The fact that the matching is improving over time supports the mechanism discussed in the paper. Headhunters provide better matches both for the firms and for the CEOs increasing the firm size and the CEO compensation.

Of course, there are shortcomings in this empirical specification because we don't know which CEOs are hired through a headhunter and which CEOs come from internal promotion or other channels. To address this issue directly one needs to collect the data on the origins of the CEO and the way she was hired. Such study would be able to analyze the difference between CEO compensation for a CEO coming through headhunters and not. Most importantly, it would be also able to determine the effect of a CEO hired by a headhunter on the firm performance. However, to collect such data set would require a tremendous amount of work and time. Also, it wouldn't be still guaranteed that such a data set would be full, because not all companies changing CEOs state whether they hired the CEO with a help from a headhunter or in a different way (for example through referrals). Another shortcoming in this analysis is that the firm size may not be a good proxy for the firm performance, other measures should be used to account for the current performance of a firm, such as cash flows or profits.

To try to overcome the lack of data on identities of the CEOs hired by headhunters I use non-compete enforceability index as an instrument for the probability to be hired by a headhunter. In the states with a high NCEI activity of headhunters is limited and therefore very few positions are filled by headhunters. The results show, indeed, that in the states with low NCEI the increase in CEO compensation after a CEO change is larger. This suggests that in the states with low NCEI more CEOs are hired by headhunters, so the improvement in matching is stronger and it leads to higher compensation.

Other studies discussing the increase in CEO compensation over the past decades offer various explanations of this phenomena. Gabaix and Landier (2008) and Gabaix et al. (2014) argue that the CEO pay increases because the average company size is increasing. Murphy and Sandino (2010) argue that the CEOs may better extract the rent from the company by hiring external compensation consultants that follow their interest. Murphy and Zabojnik (2004) show that the nature of CEO skills required to successfully run a company is changing over time and therefore more firms hire the CEOs from outside of the firm and have to pay her more.

In another paper Murphy and Zabojnik (2007) provide empirical evidence on the CEO origins at the moment of her appointment, i.e. whether she is coming from within the company or from outside, and the effect of the origin on the compensation. They study the S&P 500 companies (the largest 500 firms) during the period from 1970 to 2005. They show that during the 1970s and 1980s only 15% and 17% of CEO appointments account for the outside hires, while it increased to 26% in 1990s and almost 32.7% in 2000s. Even more importantly, they show that the external CEO receives 14.2% higher compensation on average over the full sample, with the difference being just 6% in the1970s, 15,9% in 1980s and 19.6% in 1990s. Not only the companies rely more and more on the outside CEOs but also the pay difference between the internal and external CEOs is increasing. After reconciling these results with the data that almost all of the outside CEOs are hired by headhunters, this is a strong evidence for the mechanism proposed in this paper.

Among other studies more closely related to the mechanism studied in this paper, Garmaise (2009) shows that tougher non-compete agreements regulation reduces CEO turnover and compensation. Again, this suggests that in the states with higher NCEI the activity of headhunters is limited[18] and therefore it reduces opportunities of CEOs to transit between firms and improve the efficiency of matching, therefore limiting compensation. Pan (2015) shows the importance of assortative matching between CEOs and firms for determination of the CEO compensation and the firm's performance. However, Pan (2015) doesn't consider the change in the degree of assortative matching over time or geographical differences.

7 Conclusion

This paper introduces the headhunter channel to the standard model of random matching. The fact that headhunters have better information about a worker's skill level and that they can approach workers who are not actively searching for a (new) job at this moment allows for better screening of workers and reduces labor market frictions at the top part of the wage distribution. Moreover, headhunters separate the labor market for high and low-productive firms allowing the high-productive firms to access only the high-skilled workers. Because of worker skill and firm productivity complementarities, the wages of workers hired through headhunters increase more than proportionally to the rest of the workers. Thus, the presence of headhunters generates a fat tail of the wage distribution with a larger wage share of the top 1% and 10% workers.

Quantitative analysis of the model uses a calibrated version of the model to show that introduction of the headhunter channel in otherwise standard random matching model accounts for 70% of the increase in top 10% share of wages and 40% of the increase of top 1% share of wages in the U.S. between 1970 and 2010. The results are robust to the choice of targets related to the headhunter channel. The main effect comes from the improvement in the assortative matching between workers and firms, especially at the top. The pattern and the amplitude of the improvement are comparable to the empirical estimates of the change in assortative matching in the U.S. over the same period. The headhunter channel helps to generate the strong non-linearity in the pattern of matching observed in the data.

The paper also provides the empirical evidence of the joint increase of the use of headhunters by firms and the top income shares. The paper uses cross-country data on headhunter revenues and number of hires through headhunters together with the top income shares to show that normalized hires by headhunters are a good predictor of the future growth of the top income

[18]Garmaise (2009) does not talk about headhunters in his study.

shares in European countries. Then, it also shows that the new CEOs in the U.S. get higher compensations comparing to the previous CEOs in the same companies and this effect is weaker in the states with high non-compete enforceability index, i.e., in the states that potentially limit the activity of headhunters.

References

Acemoglu, D. (2002). Technical change, inequality, and the labor market. *Journal of economic literature*, *40*(1), 7–72.

AESC. (2015). *Insights q4 2015 state of the executive search industry* (Tech. Rep.). Association of Executive Search and Leadership Consultants.

Alvaredo, F., Atkinson, A. B., Piketty, T., & Saez, E. (2013). The top 1 percent in international and historical perspective. *The Journal of Economic Perspectives*, *27*(3), 3–20.

Aoki, S., & Nirei, M. (2015). Zipf's law, pareto's law, and the evolution of top incomes in the us. *Pareto's Law, and the Evolution of Top Incomes in the US (February 12, 2015)*.

Bagger, J., & Lentz, R. (2014). *An empirical model of wage dispersion with sorting* (Tech. Rep.). National Bureau of Economic Research.

Bagger, J., Sørensen, K. L., & Vejlin, R. (2013). Wage sorting trends. *Economics Letters*, *118*(1), 63–67.

Bell, B. D., & Reenen, J. V. (2013). Extreme wage inequality: pay at the very top. *The American Economic Review*, *103*(3), 153–157.

Cappelli, P., & Hamori, M. (2013). *Who says yes when the headhunter calls? understanding executive job search behavior* (Tech. Rep.). National Bureau of Economic Research.

Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality*. *Quarterly Journal of Economics*, *128*(3).

Eeckhout, J., & Kircher, P. (2011). Identifying sorting - in theory. *The Review of Economic Studies*, rdq034.

Fama, E. F., & French, K. R. (1997). Industry costs of equity. *Journal of financial economics*, *43*(2), 153–193.

Gabaix, X., & Landier, A. (2008). Why has ceo pay increased so much?*. *The Quarterly journal of economics*, *123*(1), 49–100.

Gabaix, X., Landier, A., & Sauvagnat, J. (2014). Ceo pay and firm size: an update after the crisis. *The Economic Journal*, *124*(574), F40–F59.

Gabaix, X., Lasry, J.-M., Lions, P.-L., & Moll, B. (in press). The dynamics of inequality. *Econometrica*.

Garmaise, M. J. (2009). Ties that truly bind: Noncompetition agreements, executive compensation, and firm investment. *Journal of Law, Economics, and Organization*.

Gautier, P. A., & Teulings, C. N. (2015). Sorting and the output loss due to search frictions. *Journal of the European Economic Association*, *13*(6), 1136–1166.

Hagedorn, M., Law, T. H., & Manovskii, I. (2012). *Identifying equilibrium models of labor market sorting* (Tech. Rep.). National Bureau of Economic Research.

Håkanson, C., Lindqvist, E., & Vlachos, J. (2015). *Firms and skills: the evolution of worker sorting* (Tech. Rep.). Stockholm University, Department of Economics.

IACPR. (2003). (Tech. Rep.). International Association of Corporate and Professional Recruiters.

Jenn, N. G. (1995). *Executive search in the americas: choosing and using a headhunter*.

Jenn, N. G. (1999). *Executive search in europe: choosing and using a headhunter*.

Jones, C. I. (2015). Pareto and piketty: The macroeconomics of top income and wealth inequality. *The Journal of Economic Perspectives*, *29*(1), 29–46.

Jones, C. I., & Kim, J. (2014). *A schumpeterian model of top income inequality* (Tech. Rep.). National Bureau of Economic Research.

Katz, L. F., Kearney, M. S., et al. (2006). The polarization of the us labor market. *American Economic Review*, *96*(2), 189–194.

Lemieux, T., MacLeod, W. B., Parent, D., et al. (2009). Performance pay and wage inequality. *The Quarterly Journal of Economics*, *124*(1), 1–49.

Lise, J., Meghir, C., & Robin, J.-M. (2016). Matching, sorting and wages. *Review of Economic Dynamics*, *19*, 63–87.

Mortensen, D. T., & Pissarides, C. A. (1994). Job creation and job destruction in the theory of unemployment. *The review of economic studies*, *61*(3), 397–415.

Murphy, K. J., & Sandino, T. (2010). Executive pay and "independent" compensation consultants. *Journal of Accounting and Economics*, *49*(3), 247–262.

Murphy, K. J., & Zabojnik, J. (2004). Ceo pay and appointments: A market-based explanation for recent trends. *The American Economic Review*, *94*(2), 192–196.

Murphy, K. J., & Zabojnik, J. (2007). Managerial capital and the market for ceos. *Available at SSRN 984376*.

Pan, Y. (2015). The determinants and impact of executive-firm matches. *Management Science*.

Piketty, T. (2014). *Capital in the twenty-first century*. Harvard University Press.

Postel-Vinay, F., & Robin, J.-M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, *70*(6), 2295–2350.

Schulz, B., & Lochner, B. (2016). Labor market sorting in germany.

Shimer, R., & Smith, L. (2000). Assortative matching and search. *Econometrica*, *68*(2), 343–369.

Song, J., Price, D. J., Guvenen, F., Bloom, N., & von Wachter, T. (2016). Firming up inequality.

Teulings, C. N., & Gautier, P. A. (2004). The right man for the job. *The Review of Economic Studies*, *71*(2), 553–580.

Uren, L., & Virag, G. (2011). Skill requirements, search frictions, and wage inequality*. *International Economic Review*, *52*(2), 379–406.

A Appendix

A.1 Non-monetary costs of headhunters

In this case, it is more convenient to define two measures for firms with an open position - $G_V(p)$ for the firms using vacancy channel, and $G_H(p)$ for the firms using the headhunter channel.

Workers

Now the value functions are the following. For low-skilled unemployed workers:

$$S_U(e) = S_{UV}(e) \equiv f_V(u_V, a_V, v) \int_{\underline{p}}^{\overline{p}} (W(e,p) - U(e))\, dG_V(p) - c_{wV}.$$

For high-skilled unemployed workers:

$$\begin{aligned} S_U(e) = S_{UVH}(e) \equiv \quad & f_H(u_H, a_H, h)\,(1 - f_V(u_V, a_V, v)) \left(\int_{\hat{p}}^{\overline{p}} (W(e,p) - U(e))\, dG_H(p) - c_{wH}\right) + \\ & + f_V(u_V, a_V, v)\,(1 - f_H(u_H, a_H, h)) \int_{\underline{p}}^{\overline{p}} (W(e,p) - U(e))\, dG_V(p) + \\ & + f_H(u_H, a_H, h)\, f_V(u_V, a_V, v) \cdot \\ & \cdot \left(\int_{\hat{p}}^{\overline{p}} \int_{\underline{p}}^{\overline{p}} (\max\{W(e,p), W(e,p')\} - U(e))\, dG_H(p)\, dG_V(p') - c_{wH}\right) - c_{wV}. \end{aligned}$$

For low skilled employed workers:

$$S_{EV}\left(e,p\right) \equiv f_V\left(u_V,a_V,v\right)\int_{p}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_V\left(p'\right)-c_{wV}.$$

For high-skilled employed workers:

$$S_{EH}\left(e,p\right) \equiv f_H\left(u_H,a_H,h\right)\left(\int_{\max\{\hat{p};p\}}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_H\left(p'\right)-c_{wH}\right),$$

and:

$$\begin{aligned} S_{EVH}\left(e,p\right) \equiv\ & f_H\left(u_H,a_H,h\right)\left(1-f_V\left(u_V,a_V,v\right)\right)\left(\int_{\max\{\hat{p};p\}}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_H\left(p'\right)-c_{wV}\right)+ \\ & +f_V\left(u_V,a_V,v\right)\left(1-f_H\left(u_H,a_H,h\right)\right)\int_{p}^{\overline{p}}\left(W\left(e,p'\right)-W\left(e,p\right)\right)dG_V\left(p'\right)+ \\ & +f_H\left(u_H,a_H,h\right)f_V\left(u_V,a_V,v\right)\cdot \\ & \cdot\left(\int_{\hat{p}}^{\overline{p}}\int_{\underline{p}}^{\overline{p}}\left(\max\left\{\max\left\{W\left(e,p''\right),W\left(e,p'\right)\right\}-W\left(e,p\right);0\right\}\right)dG_H\left(p'\right)dG_V\left(p''\right)-c_{wH}\right)-c_{wV} \end{aligned}$$

Firms

The value function of firms posting a vacancy in this case is:

$$\begin{aligned} V_V\left(p\right) =\ & -c_{fV}\cdot p+\beta\Bigg(V\left(p\right)+q_V\left(u_V,a_V,v\right)\Bigg(\frac{u_V}{u_V+a_V}\int_{\underline{e}}^{\hat{e}}\left(J\left(p,e\right)-V\left(p,c'_{fN}\right)\right)dU\left(e\right)+ \\ & +\frac{u_V}{u_V+a_V}\left(1-f_H\left(u_H,a_H,h\right)\right)\int_{\hat{e}}^{\overline{e}}\left(J\left(p,e\right)-V\left(p,c'_{fN}\right)\right)dU\left(e\right)+ \\ & +\frac{u_V}{u_V+a_V}f_H\left(u_H,a_H,h\right)\frac{G_H(p)}{G_H(\overline{p})}\int_{\hat{e}}^{\overline{e}}\left(J\left(p,e\right)-V\left(p,c'_{fN}\right)\right)dU\left(e\right)+ \\ & +\frac{a_V}{u_V+a_V}\int_{\underline{e}}^{\hat{e}}\frac{\Lambda_V(e,p)}{\Lambda_V(e,\overline{p})}\left(J\left(p,e\right)-V\left(p,c'_{fN}\right)\right)dL_V\left(e\right)+ \\ & +\frac{a_V}{u_V+a_V}\left(1-f_H\left(u_H,a_H,h\right)\right)\int_{\hat{e}}^{\overline{e}}\frac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}\left(J\left(p,e\right)-V\left(p,c'_{fN}\right)\right)dL_{VH}\left(e\right)+ \\ & +\frac{a_V}{u_V+a_V}f_H\left(u_H,a_H,h\right)\frac{G_H(p)}{G_H(\overline{p})}\int_{\hat{e}}^{\overline{e}}\frac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})}\left(J\left(p,e\right)-V\left(p,c'_{fN}\right)\right)dL_{VH}\left(e\right)\Bigg)\Bigg). \end{aligned}$$

The value function of firms using headhunters is:

$$
\begin{aligned}
V_H(p) = \; & -c_{fH} \cdot p + \\
& +\beta \Bigg(V(p) + q_H(u_H, a_H, h) \Bigg(\tfrac{u_H}{u_H + a_H} \left(1 - f_V(u_V, a_V, v)\right) \textstyle\int_{\hat{e}}^{\overline{e}} \left(J(p,e) - V\left(p, c'_{fN}\right)\right) dU(e) + \\
& + \tfrac{u_H}{u_H + a_H} f_V(u_V, a_V, v) \tfrac{G_V(p)}{G_V(\overline{p})} \textstyle\int_{\hat{e}}^{\overline{e}} \left(J(p,e) - V\left(p, c'_{fN}\right)\right) dU(e) + \\
& + \tfrac{a_H}{u_H + a_H} \textstyle\int_{\hat{e}}^{\overline{e}} \tfrac{\Lambda_H(e,p)}{\Lambda_H(e,\overline{p})} \left(J(p,e) - V\left(p, c'_{fN}\right)\right) dL_H(e) \\
& + \tfrac{a_H}{u_H + a_H} \left(1 - f_V(u_V, a_V, v)\right) \textstyle\int_{\hat{e}}^{\overline{e}} \tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})} \left(J(p,e) - V\left(p, c'_{fN}\right)\right) dL_{VH}(e) + \\
& + \tfrac{a_H}{u_H + a_H} f_V(u_V, a_V, v) \tfrac{G_V(p)}{G_V(\overline{p})} \textstyle\int_{\hat{e}}^{\overline{e}} \tfrac{\Lambda_{VH}(e,p)}{\Lambda_{VH}(e,\overline{p})} \left(J(p,e) - V\left(p, c'_{fN}\right)\right) dL_{VH}(e) \Bigg) \Bigg).
\end{aligned}
$$

And the value of an open position is:

$$
\tilde{V}\left(p, c_{fN}\right) = \max\left\{ V_V(p) ; V_H(p) - c_{fN} \right\}.
$$

The quit rate is:

$$
s_Q(e, p, \omega) = \begin{cases}
f_V(u_V, a_V, v) \left(\frac{G_V(\overline{p}) - G_V(p)}{G_V(\overline{p})} \right) & if \;\; p < \tilde{p}_V(e) \;\; and \; e < \underline{e} \\
f_H(u_H, a_H, h) \left(\frac{G_H(\overline{p}) - G_H(p)}{G_H(\overline{p})} \right) & if \;\; \tilde{p}_{VH}(e) < p < \tilde{p}_H(e) \;\; and \; e \geq \underline{e} \\
\left(1 - f_V(u_V, a_V, v)\right) \cdot & if \;\; p < \tilde{p}_{VH}(e) \;\; and \; e \geq \underline{e} \\
\cdot f_H(u_H, a_H, h) \left(\frac{G_H(\overline{p}) - G_H(p)}{G_H(\overline{p})} \right) + & \\
+ \left(1 - f_H(u_H, a_H, h)\right) \cdot & \\
\cdot f_V(u_V, a_V, v) \left(\frac{G_V(\overline{p}) - G_V(p)}{G_V(\overline{p})} \right) + & \\
+ f_V(u_V, a_V, v) f_H(u_H, a_H, h) \cdot & \\
\cdot \left(1 - \frac{G_V(p)}{G_V(\overline{p})} \frac{G_H(p)}{G_H(\overline{p})} \right) & \\
0 & otherwise.
\end{cases}
$$

Aggregation

The number of firms using the vacancy channel:

$$
v = \int_{\underline{p}}^{\hat{p}} 1 dG_V(p).
$$

And the number of firms using the headhunter channel:

$$h = \int_{\hat{p}}^{\overline{p}} 1 dG_H(p).$$

And the number of searching workers is determined as before.

Balance

The aggregate balance equation is as before:

$$\phi(e,p)\left(s + s_Q(e,p)(1-s)\right) = i_E(e,p) + i_U(e,p),$$

while the inflows now are:

$$i_U(e,p) = \begin{cases} f_V(u_V,a_V,v)\frac{g_V(p)}{v}u(e) & \text{if } e < \hat{e} \\ f_H(u_H,a_H,h)(1-f_V(u_V,a_V,v))\frac{g_H(p)}{h}u(e) + & \text{if } e \geq \hat{e} \\ +(1-f_H(u_H,a_H,h))f_V(u_V,a_V,v)\frac{g_V(p)}{v}u(e) + & \\ +f_H(u_H,a_H,h)f_V(u_V,a_V,v)\left(\frac{g_V(p)}{v}\frac{G_H(p)}{G_H(\overline{p})} + \frac{g_H(p)}{h}\frac{G_V(p)}{G_V(\overline{p})}\right)u(e) & \end{cases}$$

and:

$$i_E(e,p) = \begin{cases} f_V(u_V,a_V,v)\frac{g_V(p)}{v}\int_{\underline{p}}^{\min\{p,\tilde{p}_V(e)\}}\phi(e,p')dp' & \text{if } e < \hat{e} \\ f_H(u_H,a_H,h)\frac{g_H(p)}{h}\int_{\min\{p,\tilde{p}_{VH}(e)\}}^{\min\{p,\tilde{p}_H(e)\}}\phi(e,p')dp' + & \text{if } e \geq \hat{e} \\ +f_H(u_H,a_H,h)(1-f_V(u_V,a_V,v))\frac{g_H(p)}{h}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')dp' + & \\ +(1-f_H(u_H,a_H,h))f_V(u_V,a_V,v)\frac{g_V(p)}{v}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')dp' + & \\ +f_H(u_H,a_H,h)f_V(u_V,a_V,v)\frac{g_V(p)}{v}\frac{G_H(p)}{G_H(\overline{p})}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')dp' + & \\ +f_H(u_H,a_H,h)f_V(u_V,a_V,v)\frac{g_H(p)}{h}\frac{G_V(p)}{G_V(\overline{p})}\int_{\underline{p}}^{\min\{p,\tilde{p}_{VH}(e)\}}\phi(e,p')dp' & \end{cases}$$

A.2 Wage bargaining

In the baseline model wages are set as a constant share of the production, in this extension the wages are determined in period by period wage bargaining between the worker and the firm. This might change the implication of the model because headhunters will affect the outside options of both parties. They improve the value of the vacancy for the firm, so improving firm's bargaining position and driving the wages of top earners down, potentially dampening the effect from better matching. But at the same time, they facilitate job search for high-skilled workers improving also their bargaining position and increasing their wages even more. Moreover, the bargaining position of medium-skilled workers worsens because they lose the possibility to move to better matches, therefore decreasing their wages.

As in standard Nash bargaining, wage in a match between a worker with skill e and a firm with productivity p is a solution of the Nash bargaining problem:

$$w(e,p) = \max_{w} \left(\hat{W}(e,p,w) - U(e)\right)^{\gamma} \left(\hat{J}(e,p,w) - V(p)\right)^{1-\gamma},$$

where γ is the bargaining power of the worker.

The FOC:

$$\gamma \left(\hat{W}(e,p,w) - U(e)\right)^{\gamma-1} \left(\hat{J}(e,p,w) - V(p)\right)^{1-\gamma} \frac{\partial \hat{W}(e,p,w)}{\partial w} =$$
$$- (1-\gamma) \left(\hat{W}(e,p,w) - U(e)\right)^{\gamma} \left(\hat{J}(e,p,w) - V(p)\right)^{-\gamma} \frac{\partial \hat{J}(e,p,w)}{\partial w},$$

or simply

$$\gamma \left(\hat{J}(e,p,w) - V(p)\right) \frac{\partial \hat{W}(e,p,w)}{\partial w} = -(1-\gamma) \left(\hat{W}(e,p,w) - U(e)\right) \frac{\partial \hat{J}(e,p,w)}{\partial w}.$$

From the value functions we can find that:

$$\frac{\partial \hat{W}(e,p,w)}{\partial w} = -\frac{\partial \hat{J}(e,p,w)}{\partial w} = 1,$$

so the equilibrium wage for every match must satisfy the standard sharing rule:

$$\gamma \left(\hat{J}(e,p,w) - V(p)\right) = (1-\gamma) \left(\hat{W}(e,p,w) - U(e)\right).$$

Start with the model without headhunters. RHS of the sharing rule can be written as:

$$\gamma \left(\mathbf{y} - \mathbf{w} + \beta \left(\left(s + \mathbf{s_Q} (1-s)\right) V' + (1 - \mathbf{s_Q})(1-s) \mathbf{J}'\right) - \right.$$
$$\left. - \left(-c_{fV} \cdot p + \beta \left(V' + q_V E_{e'|V} \left[P(A)(J' - V')\right]\right)\right)\right),$$

and the LHS can be written as:

$$(1-\gamma) \left(\mathbf{w} + \beta \left(sU' + (1-s)\left(\mathbf{W}' + \mathbf{S_E'}\right)\right) - \left(b + \beta \left(U' + S_U'\right)\right)\right).$$

If the worker does not search on-the-job, the expressions simplify. For the RHS:

$$\gamma\left(\mathbf{y}-\mathbf{w}+\beta\left(sV'+(1-s)\,\mathbf{J}'\right)-\left(-c_{fV}\cdot p+\beta\left(\left(1-q_V\right)V'+q_V E_{e'|V}J'\right)\right)\right),$$

and for the LHS:

$$(1-\gamma)\left(\mathbf{w}+\beta\left(sU'+(1-s)\,\mathbf{W}'\right)-\left(b+\beta\left(U'+S'_U\right)\right)\right).$$

We can solve for $\mathbf{w}$ and apply the sharing rule for the next period to get:

$$\mathbf{w}=\gamma\left(\mathbf{y}+c_{fV}\cdot p\right)+(1-\gamma)\,b+\beta\gamma\left(q_V V'-q_V E_{e'|V}J'\right)+(1-\gamma)\,\beta S'_U.$$

Without worker/firm heterogeneity this expression collapses to the standard wage equation - equilibrium value of tomorrow search will be equal to equilibrium value of a job, that in turn will be equal to the expected cost of a vacancy posted (κ/q).

Now consider the case when the worker searches on-the-job. We can solve for the wage, $\mathbf{w}$, from the initial sharing rule, applying the sharing rule of the next period when needed to obtain the following expression for the wage:

$$\begin{aligned}\mathbf{w}= \;& \gamma\left(\mathbf{y}+c_{fV}\cdot p\right)+(1-\gamma)\,b-(1-\gamma)\,\beta\left((1-s)\,\mathbf{S}'_{\mathbf{E}}-S'_U\right)\\ & -\beta\gamma\left(\mathbf{s}_{\mathbf{Q}}\,(1-s)\left(\mathbf{J}'-V'\right)+q_V E_{e'|V}\left[P\left(A\right)\left(J'-V'\right)\right]\right).\end{aligned}$$

This expression doesn't change in the case of the model with the headhunter channel (except the expectation operator). What changes with the headhunters are the values of the search for the worker, both from unemployment and employment, the value of a vacancy for the firm, and the quit rate. Effects of headhunters on wages are heterogeneous across different matches and depend dramatically on the bargaining power. For example, for the match between the top-ranked worker and the top-ranked firm, where the worker doesn't search on-the-job and the quit rate is equal to 0, the headhunter channel increases both, the outside option of the worker, S'_U, and the outside option of the firm. They have opposite effects on the wage, and which one will be stronger depends fully on the bargaining power. For other matches, the effect is even more complicated. On top of the opposing effects of outside options, there is also an effect on the worker's on-the-job search. With headhunters, the worker doesn't lose the possibility to continue search on-the-job and receive offers from better firms. This puts downward pressure on wages because the worker agrees to the match easier. Moreover, there is an interaction between the worker's search and the value of a vacant position for the firm through the quit rate. The value of a vacant position increases with headhunters, putting upward pressure on wages, but because quit rate increases at the same time, this effect is decreased leaving the overall effect ambiguous.

Numerical simulations show that the overall effect on individual wages, and, especially, on the wage distribution is ambiguous and depends crucially on the choice of the bargaining power. With a high bargaining power of the worker, the effect of headhunters on top wages is higher than in the benchmark model, while with a very low bargaining power the effect is even the opposite, with headhunters reducing wage inequality (even though the value of the bargaining power is not realistic in such simulations). When bargaining power is set to the levels used in the literature, the overall effect is close to the benchmark results.

A.3 Headhunters as profit maximizers

In this section, I extend the model to add headhunters as additional agents choosing the fee and the screening standards in order to maximize the profits. To choose the screening standard, the headhunters need to compare the expected payoff from firms willing to use headhunters with the given standard to the cost of screening. Headhunters have correct expectations about the number and the productivity of firms that will use headhunters with each screening standard. The headhunters solve the following problem:

$$\max_{\hat{e},c_{fH}} \left[\int_{\hat{p}\left(\hat{e},c_{fV}\right)}^{\bar{p}} c_{fV} \cdot p dG\left(p\right) - \int_{\hat{p}\left(\hat{e},c_{fV}\right)}^{\bar{p}} c_{H}\left(\hat{e}\right) dG\left(p\right) \right],$$

where the first part is the fee revenues from firms using headhunters, and the second term is the total cost of screening the workers. Headhunter balance between the fee and the screening standard. When the screening standard is very high, many firms will want to participate and pay a high fee for it, but the cost of screening for headhunters will be also high, reducing the profits. And when the screening standard is low, firms' willingness to use headhunters decreases, so the headhunter has to reduce the fee, and the profits drop. Solution to this problem crucially depends on the form of the screening cost function.

Modeling headhunters explicitly and calibrating the cost function to match the screening standard and the optimal fee would be equivalent to directly calibrating the standard and the fee, as in the benchmark experiments of this paper. This would change, however, if we studied a dynamic version of the model, but this is the question for future research. Another issue with modeling headhunters is the choice of the market structure. Is it a competitive market, monopoly, or monopolistically competitive market? This question is also left for the future research.